



CONTENTS

ABBREVIATIONS

LIST OF ORIGINAL PUBLICATIONS

REVIEW OF THE LITERATURE

**Szeged University Ph.D. Dissertation****TRANSCRIPTIONAL REGULATION OF THE  
MATRILIN-1 GENE**

Identification of binding sites for cartilage-specific transcription factors  
in the short promoter of the matrilin-1 gene

**Otgonchimeg Rentsendorj****Supervisor: Ibolya Kiss, Ph.D, D.Sc.**

Molecular and Cell Biology Ph.D. Program  
Faculty of Natural Science  
University of Szeged  
Hungary

Institute of Biochemistry  
Biological Research Center  
Hungarian Academy of Sciences

**Szeged, 2006**

# CONTENTS

ABBREVIATIONS	3
LIST OF ORIGINAL PUBLICATIONS	5
REVIEW OF THE LITERATURE	6
I. Genetic and molecular understanding of skeleton biology and its importance	6
Ia. Structure and importance of the cartilage	6
Ib. Chondrogenesis and growth plate, their function in the endochondral bone development	7
Ic. Role of Sox9 and Sox5/6 transcription factors in the chondrogenesis	11
Id. Other regulatory mechanisms of endochondral bone formation	12
Ie. Matrilin-1, structure, expression and role in the cartilaginous ECM	14
II. Transcriptional Regulation of Eukaryotic Genes	16
IIa. Core promoters and the general transcription machinery	16
IIb. Activated transcription	17
IIc. A general model for regulation of a gene	19
IId. HMGB proteins: dynamic players in gene regulation	20
SOX HMG-boxes	21
IIe. Transcriptional regulation of cartilage protein genes	22
IIf. Transcriptional regulation of the matrilin-1 gene: what is known?	23
AIMS OF THE STUDY	26
MATERIALS AND METHODS	27
Oligonucleotide primers	27
Chemicals and enzymes	27
Bacterial strains used for cloning or protein expression	28
Plasmid vectors	28
Culture media for eukaryotic cells	28
Cell culture preparations	28
Sequence analysis	29
Isolation of genomic DNA from tissue culture	29
<i>In vivo</i> footprinting with dimethyl sulphate and UV light	29
Preparation of cell extracts	31

Synthesis and purification of bacterially expressed GST-SOX9	31
EMSA and supershift experiments	31
Luciferase reporter gene constructs for transient expression assay	32
Site directed mutagenesis	32
Transient expression assay	33
RESULTS	34
Sequence analysis reveals several Sox-binding sites within the short promoter of the chicken matrilin-1 gene	34
<i>In vivo</i> footprinting implicate cartilage-specific occupancy of conserved Pe1 and Ine elements by nuclear proteins	35
Sox9, L-Sox5 and Sox6 interact with the Pe1 element <i>in vitro</i>	39
Mutations in either of the paired Sox motifs of Pe1 interfere with protein binding in EMSA	40
The initiator element Ine can also interact with Sox factors <i>in vitro</i>	43
Functional analysis of Pe1 and Ine elements in transient expression studies	46
DISCUSSION	50
SUMMARY OF NOVEL FINDINGS	59
ACKNOWLEDGEMENTS	61
REFERENCES	63
SUMMARY OF THE THESIS (ENGLISH)	76
SUMMARY OF THE THESIS (HUNGARIAN)	83

## ABBREVIATIONS

AMP	adenosine 5'-monophosphate
BMP	bone morphogenetic protein
CBP	CREB binding protein
CD-RAP	cartilage-derived retinoic acid-sensitive protein
CEC	chicken embryo chondrocytes
CEF	chicken embryo fibroblasts
CH3	cysteine/histidine-rich domain 3
CIAP	calf intestine alkaline phosphatase
CMP	cartilage matrix protein
COMP	cartilage oligomeric matrix protein
CREB	cyclic-AMP response element binding
CYRBP1	$\alpha$ A-Crystallin binding protein 1
DMEM	Dulbecco's modified eagle's medium
DMS	dimethylsulfate
Dpe1 and Dpe2	distal elements 1 and 2
DTT	dithiothreitol
ds	double stranded
ECM	extracellular matrix
E.coli	Escherichia coli
EGF	epidermal growth factor-like
EMSA	electrophoretic mobility shift assay
$\Delta$ EF1	estrogen-inducible transcription factor delta
FCS	fetal calf serum
FGF	fibroblast growth factor
GST	glutathione S-transferase
HAM	nutrient mixture F-12
HDM	high-density mesenchyme
HMG	high-mobility group
HNF3	hepatocyte nuclear factor 3
Ihh	Indian hedgehog
Ine	initiator element
IPTG	isopropyl- $\beta$ -D-thiogalactopyranoside
LcMaf	long cell lineage-specific musculoaponeurotic fibrosarcoma
LM-PCR	ligation mediated-PCR

MATN3	matrilin-3 (human)
MED	multiple epiphyseal dysplasia
MIA	melanoma inhibitory activity
Nce	neural crest-specific element
NFI	nuclear factor-I
PAGE	polyacrylamid gel electrophoresis
Pe1	promoter element 1
Pfu	Pyrococcus furiosus DNA polymerase
PIC	preinitiation complex
PMSF	phenylmethysulphonylfluoride
PNK	polynucleotide kinase T4
PR1 and PR2	promoter positive region 1 and 2
PTHrP	parathyroid hormone-related peptide
SF-1	steroidogenic factor-1
SI and SII	silencer element I and II
Sox	sry-type HMG box protein
Sry	sex-determining region Y
TAFs	TBP-associated factors
TBP	TATA-binding protein
t	<i>in vitro</i>
UV	ultraviolet
v	<i>in vivo</i>
vWFA	von Willebrand factor A

## LIST OF ORIGINAL PUBLICATIONS

This thesis is based on the following original publications.

- I Rentsendorj, O., Nagy, A., Sinko, I., Daraba, A., Barta, E., Kiss, I. (2005) Highly conserved proximal promoter element harboring paired Sox9-binding sites contributes to the tissue- and developmental stage-specific activity of the matrilin-1 gene. *Biochem J.* **389**, 705-16
- II Karcagi, I., Rauch, T., Hiripi, L., Rentsendorj, O., Nagy, A., Bősze, Zs., Kiss, I. (2004) Functional analysis of the regulatory regions of the matrilin-1 gene in transgenic mice reveals modular arrangement of tissue-specific control elements. *Matrix Biol.* **22**, 605-618
- III Rentsendorj, O. (2002) The chondrogenic master transcription factor Sox9 binds to the regulatory region of the matrilin-1 gene. DISSERTATION SUMMARY, *Acta Biologica Szegediensis.* **46** (1-2): 45, <http://www.sci.u-szeged.hu/ABS>

## **REVIEW OF THE LITERATURE**

### **I. GENETIC AND MOLECULAR UNDERSTANDING OF SKELETON BIOLOGY AND ITS IMPORTANCE**

For a long time, the skeleton was seen as an amorphous tissue of little biological interest. But such a view ignored the large number of genetic and degenerative diseases affecting this organ. Over the past 15 years, molecular and genetic studies have modified our understanding of skeleton biology. By so doing this progress has affected our understanding of diseases and suggested in many instances new therapeutic opportunities. To emphasize the current importance of the problem, the “Decade of Bone and Joint Disease: 2000-2010” program was launched by the WHO.

#### **Ia. Structure and importance of the cartilage**

The skeleton is made of two tissues (cartilage and bone), three cell types (chondrocytes, osteoblasts and osteoclasts) and more than 200 different skeletal elements spread out throughout the body. Beyond development, the skeleton has to fulfil a series of functions about which we have little understanding in molecular terms, but which are of critical importance as they are often affected in common degenerative diseases and/or skeletal disorders.

Cartilage is a unique connective tissue that plays essential roles in vertebrate development and adulthood. Cartilage anlagen develop in the embryo before bone, and thus provide the first skeleton of the embryo. It acts as the main source of body longitudinal growth that provides structural templates and induction signals for the formation of most bones through a process called endochondral ossification [1]. Furthermore, cartilage structures that persist through life in airways, joints, and ears are essential to our breathing, articulation, locomotion, and hearing. It is therefore not surprising that cartilage malformation diseases, which account for a large proportion of birth defects in humans, can in some cases have severe consequences for a person’s expectancy and quality of life, and can lead to embryonic or perinatal lethality or life-long handicaps. To understand the genetic causes of cartilage malformation diseases, and improve prevention and therapeutic strategies, we must reach a thorough

understanding of this process and identify all genes involved, as well as their functions, interactions, and modes of regulation.

Cartilage is an elaborate network of large macromolecules, synthesised and deposited by the specialized cells, called chondrocytes. They deposit these macromolecules around themselves to form an extracellular matrix (ECM). In return, the matrix gives the direct surrounding of the cells providing nutrients, hormones and various signaling molecules and protects the cells from damage. In addition, the ECM also serves as a medium to which the cells are attached and spread on to survive.

The framework of the cartilage matrix is a collagen fiber network that is comprised primarily of collagens types II IX and XI encoded by genes *Col2a1*, *Col9a1*, *Col9a2* and *Col9a3*, *Coll1a1*, *Coll1a2*, respectively. Collagen type X is also produced in abundance but exclusively by prehypertrophic and hypertrophic chondrocytes. This collagen composition of cartilage contrasts with that of bone and most other connective tissues, which are built on a fibrillar network of collagen types I, III, and V. The cartilage collagen network entraps a highly hydrated gel of proteoglycans and glycoproteins. Aggrecan (*Agc1*) is a large, very abundant proteoglycan that is almost unique to cartilage. It forms enormous aggregates by binding to linear chains of the glycosaminoglycan hyaluronan with the help of cartilage link protein (*Crtl1*). Glycoproteins, such as matrilin-1 (*Matn1*) and cartilage oligomeric protein, and small proteoglycans, such as fibromodulin and perlecan, vary in abundance according to the types of cartilage. They collectively make up ~15% of the dry mass of growth plate cartilage matrix. In addition to being responsible for the biomechanical properties of the tissue, the cartilage matrix also significantly modulates chondrocyte differentiation and activity [2,3]. It is clear that components of the cartilage matrix itself play important roles in either modulating or maintaining the phenotype of chondrocytes and their correct organization in the growth plate [4-6].

### **Ib. Chondrogenesis and growth plate, their function in the endochondral bone development**

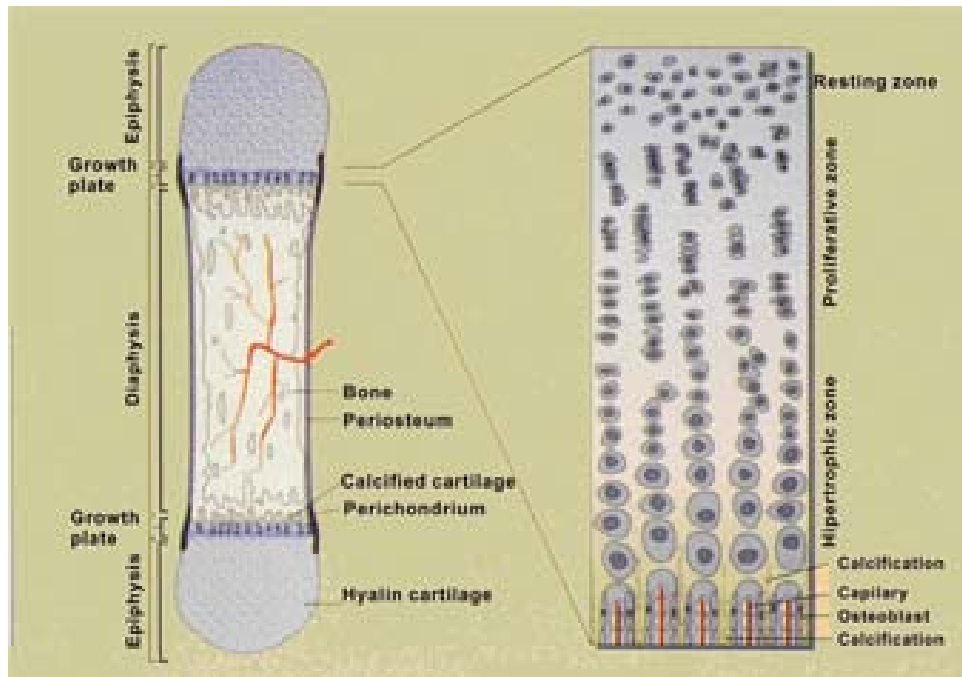
Bones form through two distinct mechanisms: intramembranous and endochondral ossification [1,7-9]. The first, in which mesenchymal cells develop directly into osteoblasts, is involved in the formation of the flat skull bones. However, most other bones form by endochondral ossification through a process involving a cartilage intermediate.



Chondrogenesis is an essential process in vertebrates. It leads to the formation of cartilage growth plates (Figure 1), which drive body growth and have primary roles in endochondral ossification. Chondrogenesis also leads to the formation of permanent cartilaginous tissues that provide major structural support in the articular joints and respiratory and auditory tracts throughout life. Defects in chondrogenesis cause chondrodysostoses and chondrodysplasias.

At the onset of skeletogenesis, mesenchymal precursor cells commit to chondrogenesis and differentiate into prechondrocytes, and then early chondroblasts that form cartilage anlagen, which prefigure future skeletal elements. The cells alter their gene expression pattern. Instead of mesenchymal markers (type I collagen, fibronectin, etc) the cells start to secrete cartilage markers (type II collagen, and aggrecan). In the center of the diaphyses of future long bones, the cells rapidly progress toward prehypertrophy, hypertrophy, terminal maturation, and ultimately undergo apoptosis. Each of these differentiation steps is characterized by the activation of a specific set of marker genes. Bone-forming cells then invade the lacunae and form primary ossification centers. On either side of these centers, early chondroblasts develop cartilage growth plates (Figure 1). Growth plate's activity lead to the continual production of new cells and cartilage matrix, thereby it plays a pivotal role in promoting longitudinal bone growth. In morphological appearance, the growth plate is unique and striking owing to its highly anisotropic and distinct cellular organization [10]. The chondrocytes are arranged into vertical columns (see Figure 1), which act as the functional units of longitudinal bone growth [10]. Cells located in the middle of the epiphyses of future long bones undergo a similar maturation to lead to the formation of secondary ossification centers, but do so several days or weeks after birth in the mouse, after a long pause at the early chondroblast stage.

The layer most distal from the diaphysis is called the reserve cell zone or resting cell zone [10,11]. It is a narrow and irregularly contoured region, consisting of single or paired chondrocytes. This zone serves two purposes in the post-fetal growth plate: it provides a mechanical buffer so that mechanical forces on the bone are appropriately dispersed, and it permits interstitial growth to occur in the growth plate proper. Cells in the resting zone are not proliferative. However, in response to signals that are only now beginning to be understood, cells at the base of the resting zone begin to proliferate and undergo a set number of cell divisions.

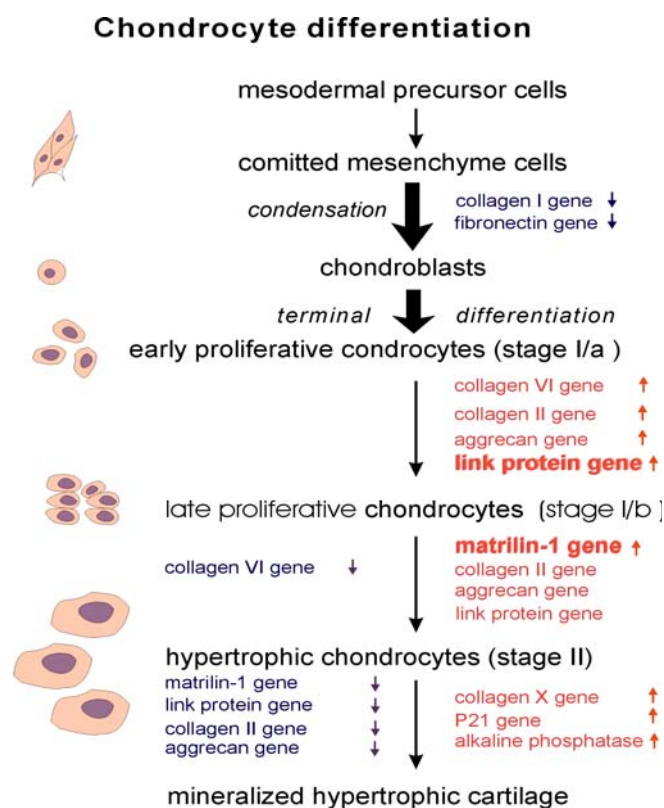


**Figure 1. Structure of a growing bone and the growth plate.** Expanded figure represents a cellular organization of the growth plate where the tissue is typically organized into vertical cell columns, each of which recapitulates the different stages: resting-, proliferating, prehypertrophic, and hypertrophic, through which a chondrocyte passes during the course of its life cycle. During the final stages of hypertrophy, chondrocytes initiate matrix calcification in the longitudinal septa, which is followed by blood vessel invasion, and finally replacement of the cartilaginous matrix by trabecular bone (the primary spongiosa).

The reserve cell zone impinges on the round proliferating one, which, as the name implies, corresponds to a phase in which chondrocytes undergo rapid division [10], thereby increasing the columnar cell pool and thus contributing to longitudinal bone growth. They proliferate at the highest rate at the top of the columns and progressively decrease their proliferation rate as they move down the columns. Cells divide perpendicular to the long axis of growth [12], but, when this process is complete, they rearrange their position to lay one above the other rather than side by side. In addition to the other cartilage protein genes, the activation of a characteristic marker gene, the matrilin-1 gene can be observed (Figure 2) [13]. After a finite number of cell divisions, proliferating chondrocytes lose their capacity to divide and begin to hypertrophy.

The early stages of hypertrophy were previously referred as the phase of maturation [10,14]. Hypertrophy contributes efficiently, and economically to longitudinal bone growth [10,15]. There is a tremendous increase in cell size (Figure 1,2) that is accompanied by rapid remodeling of the

extracellular matrix. Hypertrophic chondrocytes express predominantly type X collagen. In the lower hypertrophic cell zone, there is evidence of initial calcium phosphate crystal deposition that appears to be needed for the formation of bone marrow. In addition, hypertrophic chondrocytes modulate the formation of the bone collar. Chondrocytes undergo a last dramatic phenotypic change when they progress from hypertrophy to the terminal stage. During this transition, they activate a new set of genes, such as matrix metalloproteinase-13, and osteopontin. This process is followed by the death of hypertrophic chondrocytes, blood vessel invasion, and finally replacement of the cartilaginous matrix by trabecular bone (the primary spongiosa). The primary center splits into two opposite growth plates, in each of which the maturation of cartilage and subsequent remodeling into bone continue, as long as new chondrocytes are generated in the growth plates. As chondrocyte proliferation fuels longitudinal bone growth during post-natal life, the growth plates are separated by an increasing amount of space that becomes filled by bone marrow.



**Figure 2.** Schematic of the successive steps of the chondrocyte differentiation pathway as observed in chondrogenic culture systems *in vitro*. Major cellular features are highlighted on the left and extracellular matrix markers expressed at each differentiation steps are indicated on the right. Upregulated genes are represented in red, while downregulated ones represented in purple.

The consecutive steps of the chondrocyte differentiation pathway can be mimicked *in vitro* by culturing mesenchymal cells derived from limb buds [16]. The high-density mesenchyme culture from chicken limb buds was utilized as a model for studying the cartilage-specific ECM components at early stage I (stage I/a) of chondrogenesis. This culture was characterized by steady state mRNA levels for cartilage proteins including the later activation of the matrilin-1 gene (stage I/b) as compared to collagens type II and VI, aggrecan and link protein (see Figure 2).

As described above, molecular events lead to carefully orchestrated alterations in chondrocyte size, extracellular matrix components, secreted enzymes and growth factors, and receptor expression. Therefore, it is not surprising that any disturbance to this carefully coordinated process would result in malformation of the adult skeleton. In recent years, the tools provided by modern genetic both in mice and men have been instrumental in the process of identifying and dissecting basic molecular mechanisms of endochondral bone formation. Below, I give a brief summary of the current knowledge about some of the crucial factors involved in endochondral bone development.

### **Ic. Role of Sox9 and Sox5/6 transcription factors in the chondrogenesis**

Many lines of evidence have shown that Sox9, L-Sox5, and Sox6 proteins are necessary for chondrocyte differentiation. These Sox proteins contain a SRY-related (sexdetermining region Y gene) HMG (high-mobility-group)-box DNA-binding domain [17,18].

**Sox9** is master regulator of the chondrocyte lineage [19]. It is required for the activation of cartilage protein genes such as *Col2a1*, *Col9a1*, *Col11a2*, and *Agc1* [20,21]. The first clue that Sox9 plays a role in chondrogenesis came with the identification of heterozygous mutations in the *SOX9* gene in human patients with camptomelic dysplasia (CD), a severe form of chondrodysplasia that is often associated with XY sex reversal and malformations in several internal organs [22,23]. Sox9 transcripts are detected in all prechondrogenic mesenchymal condensations as early as 8.5 to 9.5 days of mouse embryonic development, and the expression peaks in cartilage primordia at 11.5 to 14.5 days [24]. Like collagen type II, encoded by *Col2a1*, *Sox9* is expressed at high levels in all prechondrocytes and chondrocytes. In the growth plate of long bone, *Sox9* and *Col2a1* gene expression levels are high in both resting and proliferating chondrocytes, while *Sox9* is switched off in hypertrophic chondrocytes

where *Col2a1* transcripts are still detectable. Several genetic approaches in the mouse (gain/loss of function) have demonstrated that *Sox9* positively regulates proliferation and negatively regulates chondrocyte hypertrophy [25-27].

**Sox5 and Sox6** have essential and redundant roles in early chondroblasts [28]. L-Sox5 is a longer product of the *Sox5* gene than the *Sox5* protein initially identified in adult testis, and it is highly identical to *Sox6*. L-Sox5 and *Sox6* share only partial identity with *Sox9* in the HMG-box DNA-binding domain. They have no transactivation or transrepression domains, and may thus act mainly to facilitate organization of transcriptional complexes. Based on their structure and roles *in vivo*, it is likely that their molecular roles are virtually identical, but different from those of *Sox9*. The expression of *Sox5* and *Sox6* requires *Sox9* [25]. They are activated in prechondrocytes and highly expressed in chondroblasts in all developing cartilage elements of the mouse embryo [29]. Although individual *Sox5* or *Sox6* null mice are born with minor cartilage defects, double knockout mice develop a severe, generalized chondrodysplasia characterized by a virtual absence of cartilage due to a defect in cell proliferation and impairment of cartilage matrix production [28]. This severe impairment occurred despite normal expression of *Sox9*, indicating that *Sox9* requires *Sox5* and *Sox6* to drive overt chondrogenesis. Consistent with above *in vivo* results, *in vitro* experiments have also suggested that, besides *Sox9*, L-*Sox5* and *Sox6* are also involved in *Col2a1* expression. They appear to form a large complex with each other and other nuclear proteins in chondrocytes. Therefore, it is believed that the three *Sox* genes cooperatively activate the *Col2a1* [29-31]. Moreover, the three *Sox* genes together were also able to suppress expression of markers for hypertrophic chondrocytes and osteoblasts. This experiment thus confirmed that L-*Sox5*, *Sox6*, and *Sox9* constitute a master chondrogenic trio.

#### **Id. Other regulatory mechanisms of endochondral bone formation**

Growth plate activity is subject to regulation by a number of other factors. Many hormones and growth factors have been shown to regulate chondrocyte differentiation [32]. These include bone morphogenic protein, Indian hedgehog, parathyroid hormone-related peptide, fibroblast growth factor and its ligand FGF18, and others. In the presence of a mutation in a receptor for one of the aforementioned growth factors, chondrodysplasia occurs.

*Bone morphogenetic proteins (BMPs) in endochondral bone formation.* Early in the limb, BMP signaling is known to play a crucial role in formation of mesenchymal condensations and the formation of the joints. It has been shown that the BMP antagonist Noggin suppresses the formation of mesenchymal condensations [33], whereas cartilage primordia is enlarged in *Noggin* knocked-out mice [34]. BMP-2, -4, and -7 coordinately regulate the patterning of limb elements within the condensations depending upon the temporal and spatial expression of BMP receptors and BMP antagonists, such as noggin and chordin [33,35]. It has been shown using various overexpression systems that BMPs are positive modulators of chondrocyte proliferation, and they negatively regulate chondrocyte terminal differentiation [36,37].

*Ihh-PTHrP signaling in endochondral bone formation.* Parathyroid hormone related peptide (PTHrP) is expressed in periarticular chondrocytes and this expression is dependent on Indian hedgehog (Ihh), whereas its receptor, the PTHrP receptor (PPR) is expressed at low levels in proliferating chondrocytes but at much higher levels in prehypertrophic chondrocytes. PTHrP and Ihh establish a negative feedback loop. Ihh is expressed at the prehypertrophic–hypertrophic boundary so that cells that escape the inhibitory action of PTHrP signaling in the growth plate express Ihh, which in turn will stimulate PTHrP expression [38]. Ihh appears as a coordinator of endochondral ossification, regulating chondrocyte proliferation and differentiation and osteoblast differentiation, and coupling chondrogenesis to osteogenesis [39].

*FGF signalling in endochondral bone formation.* FGF signalling crucially regulates chondrocyte proliferation and differentiation. Many of the 22 distinct *FGF* genes and four *FGF receptor (FGFR)* genes are expressed at distant steps of endochondral bone formation [40]. In addition, FGF signaling activates several signaling pathways, including the phosphorylation of Stat1, its translocation to the nucleus and an increased expression of the cell cycle inhibitor p21 [41,42]. Mutations in the *FGFR3* gene are the cause of achondroplasia, hypochondroplasia and thanatophoric dysplasia, that all three cause severe dwarfism in humans [43-47]. These mice show a marked decrease in the proliferation rate of the columnar proliferating chondrocytes and decreased size of the zone of hypertrophic chondrocytes [48-52]. Interestingly, in these mice, expression of Ihh is decreased. Thus, a normal function of FGF signaling in chondrocytes is to inhibit chondrocyte proliferation.

### **Ie. Matrilin-1, structure, expression and role in the cartilaginous ECM**

Matrilin-1/CMP (cartilage matrix protein) was the first identified member of the multidomain family of proteins implicated in the organization of the ECM [53-56]. It was first purified from a bovine tracheal cartilage [Paulsson and Heinegard 1981], and the molecular weight was reported that the bovine CMP formed a disulfide-bonded trimer with a  $M_r$  of 148,000. In 1989, members of our group isolated and characterized for the first time the entire gene for chicken matrilin-1 [54]. Furthermore, with the cloning of mouse [57] and human matrilin-2 [58], in 1997 and 2000 respectively, our group discovered the protein family called matrilins. To date, four members of the family have been identified that all share the structure made up of one or two von Willebrand factor A (vWFA) domains, variable number of epidermal growth factor-like (EGF) domains and a C-terminal coiled-coil oligomerization domain [56] (Figure 3A). Matrilin-1 and matrilin-3 are expressed mainly in hyaline cartilage, while matrilin 2 and matrilin 4 are expressed in a wide variety of extracellular matrices [57,59-62]. Yet, all matrilins are expressed in the skeletal elements during mouse limb development [63] suggesting an important function in endochondral bone formation. This is supported by a recent observation [64] showing that mutations in the exon coding for the vWFA domain of the human matrilin-3 gene lead to multiple epiphyseal dysplasia (MED, malformation of the "growing portion" or head of the long bones -epiphyses). The vWFA domains are thought to mediate interactions with other proteins, and their involvement in oligomerization, filamentous network formation, and cell adhesion and spreading has been described [65].

Previously, members of our group cloned the gene for chicken matrilin-1 [54]. Based on restriction mapping, electron microscopic experiments and nucleotide sequence analysis, the gene was determined to be 18-kb long and consist of 8 exons and 7 introns [54]. The matrilin-1 monomer consists of two vWFA domains, an EGF-like module and an  $\alpha$ -helical coiled-coil domain [53,54]. Electron microscopy, together with the sequence information, showed that the encoded protein for matrilin-1 is 54-kDa subunits assemble into compact homotrimers via their C-terminal coiled-coil  $\alpha$ -helix [55]. However, recent studies show that matrilin-1 can also form heterooligomers with other matrilins [66,67]. Seven novel matrilin oligomerization isoforms were observed [67], where matrilin-1 formed complexes with matrilin-2, -3, and -4. No interactions, however, were observed between

matrilin-2 and matrilin-3 or between matrilin-3 and matrilin-4. Hence, matrilins may play a role in stabilizing the extracellular matrix structure, since they can self-associate into supramolecular structures, resulting in the formation of filamentous networks [68-72].

Matrilin-1 is tightly associated with aggrecan [73] and type II collagen fibrils [68], the two major macromolecular networks of cartilage. It also forms collagen-independent pericellular filaments via the vWFA domains [70], and interacts with  $\alpha 1\beta 1$  integrin [74]. Therefore, it seems to perform an adapter function in the assembly of the cartilaginous matrix. Lining up with this hypothesis, inactivation of the matrilin-1 gene in transgenic mice altered the collagen fibrillogenesis [75], although, presumably due to the overlapping functions of matrilin-3 and -4 expressed at high levels in the same tissue, it did not lead to skeletal abnormalities [75,76]. Nevertheless, autoantibodies against matrilin-1 were detected in relapsing polychondritis patients [77] and increased matrilin-1 level was shown in knee osteoarthritic cartilage [78] and in rheumatoid arthritis [79].

Matrilin-1 has the most restricted expression pattern among the matrilins [56]. Apart from a few non-chondrogenic tissues [63,80], matrilin-1 is secreted only in hyaline cartilage, in variable amounts depending on the form and age of cartilage [81] and the developmental stage of chondrocytes [69,82,83]. In the cartilage primordia of the developing skeleton, the onset of the matrilin-1 gene is delayed as compared to the genes for well-known cartilage-specific markers (e.g. type II collagen, aggrecan and cartilage link protein) [84-86]. In situ hybridization experiments revealed a zonal distribution of matrilin-1 mRNA in the growth plate of long bones [87]. The highest level of gene transcript was found in the proliferative and upper hypertrophic zones, whereas the lower hypertrophic, calcified regions were negative [88]. It has been confirmed during in vitro chondrogenesis that the matrilin-1 gene is activated only in late proliferative (stage Ib) chondrocytes [82,83], as opposed to other cartilage protein genes (e.g. for collagen types II and IX, and cartilage link protein), which are turned on in early proliferative (stage Ia) chondrocytes [88]. Due to this property, the matrilin-1 gene can serve as an important developmental stage-specific marker gene for studies on endochondral bone formation and regulation of cartilage-specific gene expression [69,83].



## II. TRANSCRIPTIONAL REGULATION OF EUKARYOTIC GENES

The first, very important level of regulation of gene expression is transcriptional control. Transcriptional regulation is vastly more complex in animals than it is in microbes. In microbes with 800–6,000 genes to be regulated, their promoters are generally limited to 100–200 bp regions around the transcription start site with each promoter typically controlled by only 1–4 sequence-specific DNA binding proteins, and a given microbial species exist in only 1–3 cell or spore types. By contrast, in animals with 14,000–80,000 genes, many promoters are spread over tens of kilobases. Forty or more sequence-specific DNA binding proteins may regulate promoters in an animal, which generally has 80–250 distinct cell types. Thus, virtually every cell in an animal may have a unique combination of gene transcription pattern that is tightly regulated. The most characteristic requirement of gene control in eukaryotes is the execution of precise developmental decisions so that the right gene is activated in the right cell at the right time during development of the many different cell types that collectively form a multicellular organism.

### IIa. Core promoters and the general transcription machinery

The eukaryotic transcription takes place via an interplay of *cis*-acting DNA elements and *trans*-acting factors. The *cis* control elements include the core promoter, the promoter-proximal and the distant regulatory elements. A typical core promoter encompasses DNA sequences between approximately -40 and +50 relative to a transcription start site [89]. Core promoter DNA elements (TATA motif and initiator element) 1) bind to and control assembly of the preinitiation complex (PIC) containing RNA polymerase II (Pol II), and the general transcription factors; 2) position the transcription start site and control its directionality; and 3) via coactivators interacting with general transcription factors (GTFs), it can respond to nearby or distal activators and repressors in a cell. The core promoter alone is generally inactive *in vivo*, but *in vitro* it can bind to the general machinery and support low or “basal” levels of transcription. The TATA motif, TATAAA, is located 25–30 bp upstream of the transcription start site. The initiator element [90] directly overlaps the transcription start site.

The so-called “general transcription machinery” play a pivotal role in directing a gene transcription. Mammalian gene transcription involves a complicated interplay between activators, repressors, the general transcription machinery, and chromatin. The general transcription machinery consists of Pol II, the GTFs that include TFIIA, TFIIB, TFIID, TFIIIE, TFIIF, and TFIIF [91], and complex coactivators called mediator. Pol II is a large multisubunit enzyme. One important feature of Pol II is the heptapeptide repeat constituting carboxyl terminus of the largest subunit. This domain is phosphorylated extensively by different kinases involved in transcription regulation. The GTFs are highly conserved throughout the eukaryotes. Biochemical studies show that the GTFs support basal transcription and carry out many of the catalytic functions required for initiation.

TFIID is the only GTF capable of binding core promoter DNA both independently and specifically. It is a multisubunit protein containing TATA-binding protein (TBP) and 10 or more TBP-associated factors (TAFs). While TBP [92,93] makes direct contact with the TATA motif, the two independent proteins, TAFs and Pol II, interact with the initiator element [94].

### **IIb. Activated transcription**

The regulatory promoter is defined as the region surrounding the core promoter and within a few hundred base pairs of the transcription start site. Regulatory promoters have been defined as modulatory DNA structures, which contain a complex array of cis-acting regulatory elements required for accurate and efficient initiation of transcription of a given gene. These promoter-proximal elements can interact with ubiquitous and tissue-specific factors (activators and repressors). Lately, it is more evident that they are also prime target elements through which diversity and flexibility in the complex patterns of gene expression in multicellular organisms is created. The distant regulatory elements (enhancers and silencers) can modulate the promoter activity by interacting with activators and repressors. An enhancer is defined as a control region found at greater distance from the transcription start site, either upstream or downstream of the gene or within an intron. Recent studies demonstrated that enhancers could be linked via a protein bridge to a core promoter and still retain functionality [95]. The view is that enhancers bind activators and other sequence-specific proteins that are involved in chromatin remodelling. Once bound, these activators loop out the intervening DNA to interact with

proteins bound to the regulatory and core promoters (i.e., other activators and the general transcription machinery). These interactions believed to stabilize transcription complex assembly and facilitate long-range protein-protein interactions.

The development of multicellular organisms occurs as embryonic cells commit to specific cell lineages and take on differentiation functions. By allowing cells to modify their gene expression panel, transcription factors play key roles in both determining and effecting cell decisions. In addition, it has become clear that transcriptional activation of a given gene is defined not only by the activity of an individual factor or a single DNA-binding site, but it also depends on combinatorial interactions between multiple proteins [96]. However, how the different combinations of activators, some cell type-specific and some ubiquitous, result in cell type-specific transcription is currently not understood.

Activators are modular proteins with distinct domains for DNA binding and transcriptional activation [97,98]. The DNA-binding domain targets the activator to the major groove of the DNA at a specific sequence motif, perhaps in conjunction with cooperativity domains that allow combinatorial interactions with other activators. Regulatory proteins are often grouped into families according to the sequence and structure of their DNA-binding domains [99]. Some common motifs whose structures have been solved include the homeodomain, a variety of zinc-nucleated DNA-binding domains, the basic leucine zipper (bZIP), basic helix-loop-helix (bHLH), high-mobility-group (HMG) domain, Rel homology, Ets homology, Myb homology, and others [99]. The activation domain, on the other hand, interacts with the general machinery. The popular definition of activation domain is a region of protein that stimulates transcription when attached to a heterologous DNA-binding domain in a so-called domain swap experiment. Many activation domains have an unusually high percentage of particular amino acids. For example, Gal4, Gcn4, and most other yeast transcription factors have activation domains that are rich in aspartic and glutamic acids.

Gene expression is often regulated by repressors and corepressors. Repression mechanisms are, however, less well understood than activation mechanisms. In general, transcriptional repression can be divided into three broad categories. First, repression can occur by inactivation of an activator, which can be accomplished by several distinct mechanisms: 1) posttranslational modification of the activator [100]; 2) dimerization of the activator with non-functional partner [101]; 3) competition for

the activator's binding site or a direct repressor-activator interaction that results in masking of the activator's function [102]. Second, repression can be mediated by proteins that associate tightly with GTFs and thereby inhibit the formation of a PIC. The third category of repression is mediated by a specific DNA element and DNA-binding protein, which act dominantly to repress both activated and basal transcription of a given gene.

### **IIc. A general model for regulation of a gene**

During the process of development, genes are turned on and off in a pre-programmed fashion, a process that eventually generates cell specificity. This developmental program is orchestrated by transcription factors, which bind to specific DNA sites near genes they control. A single transcription factor is not dedicated to each regulatory event. Instead, different combinations of ubiquitous and cell-type-specific proteins are used to turn genes on and off in different regulatory context [96].

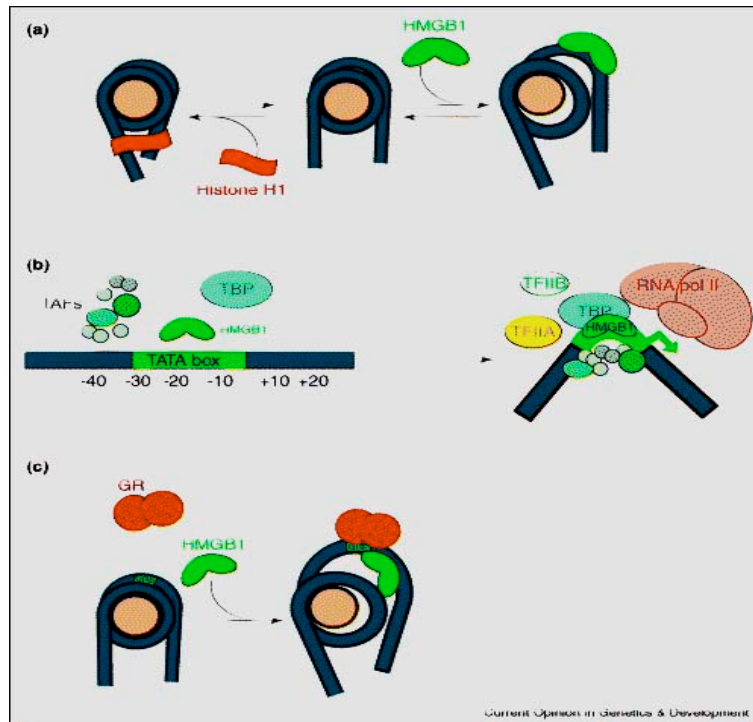
The template for transcription in eukaryotes is not free DNA, but chromatin. Chromatin maintains genes in an inactive state by restricting access to RNA polymerase and its accessory factors. The core histones, linker histones, and HMG proteins package the entire DNA in the nucleus within nucleosomal arrays [103]. Nucleosomes themselves are assembled into higher-order structures. By condensation, 2 m of DNA are compacted into about 5  $\mu\text{m}$  eukaryotic nuclei. Thus, it is necessary to open the compacted chromatin into a relative extended state to activate a gene. To overcome the nucleosome obstacle, eukaryotic cells have developed a series of enzymes (like ATP-dependent remodelling enzymes and histone acetyltransferases) to remodel the chromatin and regulate transcription. Once they bind near a gene, these enzymes remodel the chromatin so that other activators and the general machinery can bind. Binding of transcription factors is cooperative, where one protein binds weakly, but multiple factors engage in protein-protein interactions that increase each of their affinities for the regulatory region. The nucleoprotein structures comprising these combinatorial arrays of activator proteins are called enhanceosomes. The enhanceosome interacts with the general transcription machinery and recruits it to a core promoter to form the PIC. The enhanceosome, the general machinery, and the core promoter form a complicated network of protein-protein and protein-DNA interactions that dictate the frequency of transcription initiation.

In many instances genes are activated transiently and then later turned off. The mechanisms for inactivating a gene vary, but generally they involve the binding of sequence-specific repressors to silencer elements.

#### **IId. HMGB proteins: dynamic players in gene regulation**

Core histones package the genome into nucleosomes and control its accessibility to transcription factors. High mobility group proteins (HMGs) are, after histones, the second most abundant chromatin proteins and exert global genomic functions in establishing active or inactive chromatin domains. There are three families of HMG proteins (HMGA, HMGB and HMGN) with systematic reference to their structural domains [104]. HMGA proteins contain AT-hooks, exemplified by HMGI(Y) that bind AT-rich DNA stretches in the minor groove; HMGB proteins contain HMG-boxes, exemplified by HMG1 and 2 that bind into the minor groove of DNA with no sequence specificity; HMGN proteins, represented by HMG14 and 17, bind directly to nucleosomes, between the DNA spires and the histone octamer. By orchestrating multiple protein-protein and protein-DNA interactions, HMGs assist the formation of higher-order transcription factor complexes that regulate gene expression.

Mammalian HMGBs are characterized by two tandem HMG-box domains followed by a long acidic tail. HMG-box domains consist of 80-amino-acid residues that form three  $\alpha$ -helices in a twisted L-shape structure [105]. The concave surface of the L-shape binds to the minor groove of the DNA [106]. All HMG-boxes whether present alone or in tandem, are capable of bending DNA, and this might be their main function. HMGBs can promote the transcription of several genes, through several mechanisms (Figure 3). The first mechanism is the ability of HMGBs to interact directly with nucleosomes [107-110]. HMGBs can loosen the wrapped DNA and so enhance accessibility to chromatin-remodelling complexes and transcription factors. The second mechanism is the interaction of HMGBs with TBP and other GTFs [111,112]. Much better characterized is the third mode, where HMGBs establish protein-protein interaction with specific transcription factors, including all steroid receptors, HOX and POU proteins, p53 and p73, several NF- $\kappa$ B subunits [107], and SREBPs [112].



**Figure 3. Mechanisms of gene expression controlled by HMGB proteins** a) HMGBs interact with nucleosomes. HMGB1 facilitates nucleosome-remodelling. b) HMGB1 bends promoter DNA, thus increasing TBP affinity for the TATA box. The recruitment of TFIIB, TFIIA and RNA pol II follows with increased efficiency. c) HMGB1 can bend the DNA and can make the DNA sequence accessible to transcription factors, promote the recruitment of further interacting proteins. As an example, GR-binding to chromatin is stabilized by HMGB1 [107].

**SOX HMG-boxes.** SOX factors belong to the HMG-box family proteins. They contain a 79-amino-acid DNA-binding HMG-box domains that recognize 5'-(A/T)(A/T)CAA(A/T)G-3' motifs with loose sequence specificity [113-115] and regulate gene transcription. SOX factors bind to the minor groove of the DNA with lower affinity than the classical transcription factors [116]. When bound to DNA, the DNA strand exhibits 70–85° bends [117]. Why bend DNA? By altering local chromatin structure at specific binding sites, SOX proteins may act to facilitate the interaction between other factors bound at adjacent sites or allow the interaction between GTFs and nucleoprotein complexes organized on distant enhancers [118,119]. Alternatively, binding of SOX proteins and the local changes in chromatin structure may lead to the recruitment of higher-order architectural factors. Conversely, the severe distortion brought about by bending the DNA helix could act in a negative fashion by simply preventing the binding of factors to adjacent sites in the major groove.

At present, the SOX group includes at least 20 members that all display classical transcription factors and/or architectural components of chromatin. Three important features have been revealed that might be applicable to SOX HMG-box domains. First, SOX HMG-boxes share highly similar DNA-binding properties, recognizing only 6-7 bp short sequences in the minor groove, thereby loosening the hydrogen-bonds and bending the DNA strands [120,121]. Second, although a target site of a SOX-regulated gene is bound similarly by the authentic SOX and by other SOX proteins *in vitro*, only the authentic SOX protein can regulate that gene *in vivo*. Third, the same SOX protein appears to regulate different sets of target genes depending on the cell type in which the protein is expressed.

In the case of certain SOX proteins, they regulate the target genes by pairing with specific partner factors [122,123]. For example, Sox9 is known to interact with a number of partner factors, including Lc-Maf, SF-1, and HSP70 [124-126]. SOX2 is also likely to participate, along with the POU domain transcription factor OCT3/4, in the transactivation of *Fgf4* in teratocarcinoma cells [127].

The dynamic and diverse patterns of expression of *SOX* genes and analysis of mutations in humans, mice and *Drosophila* suggest that SOX factors play key roles in decisions of cell fate during diverse developmental processes. For example, SOX9 is implicated in chondrocyte differentiation [21], SOX10 in neural crest specification [128], SOX17 in endoderm specification [129], and SOX18 in endothelial cell differentiation [130].

## **Ile. Transcriptional regulation of cartilage protein genes**

Sox9 has been shown to be an activating transcription factor, indispensable for chondrogenesis [21]. The study of chimeric mice that harbor Sox9<sup>-/-</sup> cells revealed that cells without Sox9 do not express chondrocyte marker genes such as *Col2a1*, *Col9a2*, *Coll1a2*, and *Agc1* [21]. During chondrogenesis, Sox9 is expressed in chondroblasts, and subsequently, coexpressed with cartilage-specific genes, such as *Col2a1*, *Coll1a2*, *Agc1*, and *CD-RAP*.

Among the cartilage-matrix genes, transcriptional regulation of *Col2a1* is the best characterized. Role of the intronic enhancer in the cartilage-specific regulation of the gene was demonstrated [123]. Within the enhancer a 48 bp DNA element was identified, which enhanced the

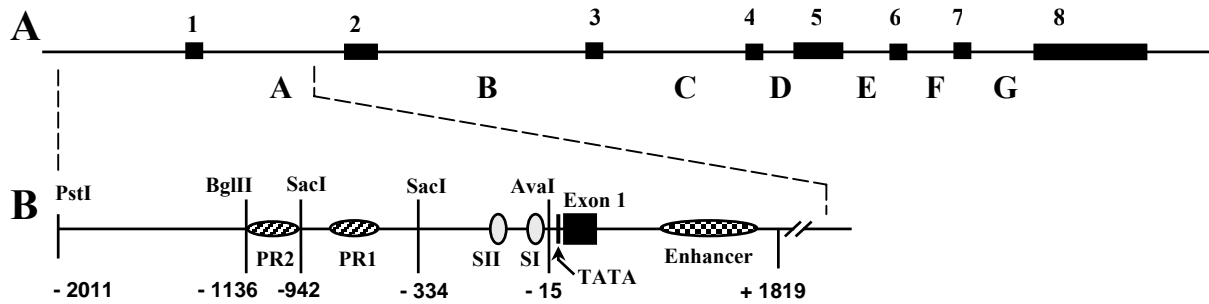
promoter activity in multiple copies in cartilage in transgenic mouse embryos [123]. The transcription factors L-Sox5, Sox6, and Sox9 bound and cooperatively activated this enhancer *in vitro* [131]. Sox9 binds to and directly activates the cartilage-specific regulatory elements of *Coll1a2* as well, which can direct cartilage-specific expression of the reporter gene in transgenic mice [31,132-134]. The similar expression pattern of these genes (*Col2a1* and *Coll1a2*), suggested that the two genes regulated in a similar manner. Sox9 also transactivates the promoter activities of *Col9a1*, *Agc1*, and *CD-RAP/MIA* [135-137]. It is not clear, however, whether the same or parallel regulatory pathways direct the cartilage-specific expression of all cartilage protein genes, and what is the molecular basis for the sequential activation and the different extraskelatal expression of these genes.

Additional Sry-related transcriptional factors, such as L-Sox5 and Sox6, are also involved in chondrogenesis. They form homo-dimers or hetero-dimers with each other and bind to the tandem repeat of HMG like sites of the tissue-specific element of *Col2a1* [29]. Sox5 and Sox6 is likely to contribute to the transcriptional activation of other cartilage protein genes, because in mice double null for Sox5 and Sox6, prechondrocytes are unable to progress along the differentiation pathway, expressing a very low level of early cartilage genes, such as *Col2a1* and *Agc1*, but they fail to turn on stage-specific matrix genes such as *COMP* and *Matn-1*.

### **III. Transcriptional regulation of matrilin-1 gene: what is known?**

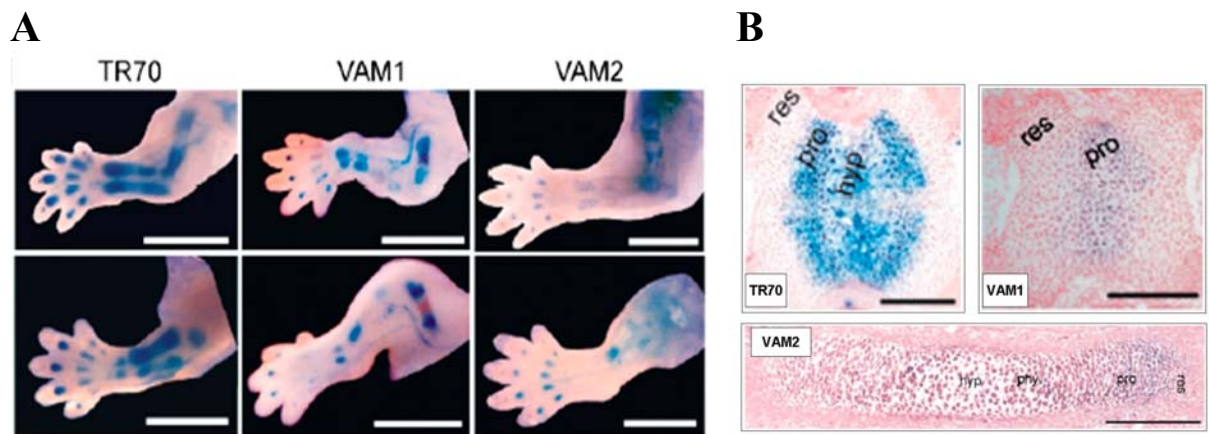
Previously, our group cloned the gene for chicken matrilin-1 [54], the first member of the matrilin family of multiadhesion proteins. The promoter of chicken matrilin-1 contains a TATA motif, TTAATA, which can function as a minimal promoter in transient expression assay [54]. The major control regions of the chicken matrilin-1 were mapped (Figure 4B) [138-140], including TATA proximal silencer elements, SI and SII, functioned by binding of NFI-family proteins. In addition, chondrocyte-specific positive control region in the first intron was found using transient expression studies [138]. Furthermore, the promoter fragment between -1137 and +64 conferred tissue- and developmental stage-specific regulation to the reporter gene as well due to an interplay between two positive (PRI and PRII) and two negative regions (SI and SII) [139].





**Figure 4. Chicken matrilin-1 gene structure** A) Physical map of the chicken matrilin-1 gene. Exons are *numbered* from 1 to 8, while introns are *labeled* from A to G sequentially. B) Known regulatory elements of the chicken matrilin-1 gene. All positions are given in bp from the first T in the TATA motif of the gene [139]. Closed, open and dotted circles represent the known regulatory elements characterized previously. Starting from the 5' end, PR2 (from -1136 to -942) and PR1 (-799 to -669) are both positive regulatory regions; SII (-145 to -120) and SI (-37 to -12) are both silencer elements function via binding of NFI nuclear proteins; an enhancer was found in the intron A, between nt +578 and +966. Well mapped restriction sites and positions are shown.

Recently, we have provided evidence in transgenic mice that the chicken long promoter (from -2011 to +67) alone in TR70 or with the intronic fragment (from -2011 to +1819) in VAM1 and the short promoter with the intronic fragment (from -338 to +1819) in VAM2 are equally able to direct the differentiation stage-specific expression of the reporter gene in chondrocytes (Figure 5) [140].



**Figure 5. Histological analysis/X-Gal staining of transgene expression in lines TR70, VAM1, and VAM2**  
**A)** Differences in the transgene expression patterns in the forelimbs (upper panels) and hindlimbs (lower panels) of transgenic embryos from lines TR70, VAM1 and VAM2. Efficient X-gal staining is visible in most of the primordial carpal and tarsal bones in line VAM1. The transgene is also stained in the developing tarsals, but only very faintly in the carpals in line TR70. Neither of these elements is stained in line VAM2. Scale bar, 2 mm.  
**B)** Transgene expression is visible in the zone of proliferative (pro) chondrocytes in the developing tibia in line VAM2, and in the proliferative, but hardly in the resting (res) and hypertrophic (hyp) chondrocytes of the developing coccygeal bones in line TR70 and VAM1. Scale bar, 200 mm.

In congruence with the expression pattern of the endogenous matrilin-1 gene, activity of the transgenes was restricted to the columnar proliferating and prehypertrophic zones of the growth plate. However, the presence of both promoter upstream and intronic elements was necessary for the high-level transgene activity in all chondrogenic tissues most resembling to that of the endogenous gene [140].

## AIMS OF THE STUDY

As described in the introduction, the matrilin-1 gene serves as a marker gene during endochondral ossification. Among the cartilage protein genes, it has a unique expression pattern, restricted to certain chondrocyte developmental stages and distinct zones of the growth plate. This suggests that the regulation of the gene may involve similar as well as different molecular mechanisms as compared to other cartilage protein genes, such as the well-characterized *Col2a1* gene. To get insight into the common and distinct molecular mechanisms controlling cartilage-specific gene expression, we aimed to study DNA elements and transcription factors, involved in the transcriptional regulation of the matrilin-1 gene.

Functional analysis of the main regulatory regions of the gene revealed that the long promoter with and without the intronic enhancer region, as well as the short promoter with the intronic enhancer region equally directed the transgene expression to distinct zones of the growth plate in transgenic mice. Based on this observation, we raised the question, whether the short promoter may have a role in the tissue-specific regulation and whether it involves cartilage-specific and/or developmental stage-specific control elements. To answer this question we aimed to focus on the characterization of the short promoter.

The following specific goals were set out:

1. to delineate *cis*-regulatory elements that are important for the transcriptional regulation of the matrilin-1 gene and to identify proteins binding to these sites, especially focusing on those sites, which possibly function via interacting with Sox-family proteins in the short promoter region.
2. to verify Sox protein-binding sites in the short promoter using *in vitro* gelshift and supershift assays.
3. to reveal tissue-specific binding of transcription factors to the short promoter region of the chicken matrilin-1 gene using *in vivo* footprinting.
4. to study the role of these putative elements in the transcriptional regulation of the matrilin-1 gene.

## MATERIALS AND METHODS

### Oligonucleotide primers

		Primer sequences used for in vivo footprinting
1	PU1	5' -TGTTCCCATCCCAGATTCC-3'
2	PU2	5' -TCCCAGATTCCCCACATACCGC-3'
3	PU3	5' -GCGTGGCTGCGGGTCCCT-3'
4	PL1	5' -GGAGGTGCCCCCAGA-3'
5	PL2	5' -TGCCCCCAGACTCCACAGCT-3'
6	PL3	5' -CCACAGCTCTGGAGGAGAAGCAG-3'
7	LinkerPeptide11	5' -GAATTCAGATC-3'
8	LinkerPeptide25	5' -GCGGTGACCCGGGAGATCTGAATTC-3'
		Double stranded (ds) oligonucleotides used for EMSA
9	Ine	5' -ACCTCAGGGCCAAGGCTCCCGTGTGCCATTCTGCATCCAACCTCC-3'
10		3' -TGGAGTCCCGGTTCGAGGGGCACACGGTAAGACGTAGGTTGGAGG-5'
11	IneM1*	5' -ACCTCAGGGC <b>gt</b> AG <b>t</b> CTCCCGTGTGCCATTCTGCATCCAACCTCC-3'
12		3' -TGGAGTCCCG <b>ca</b> T <b>ca</b> GAGGGGCACACGGTAAGACGTAGGTTGGAGG-5'
13	IneM2*	5' -ACCTCAGGGCCAAGGCTCCCG <b>g</b> C <b>g</b> G <b>t</b> Ag <b>g</b> CaGCATCCAACCTCC-3'
14		3' -TGGAGTCCCGGTTCGAGGG <b>Cg</b> C <b>g</b> Ca <b>t</b> cc <b>g</b> tCGTAGGTTGGAGG-5'
15	IneM3*	5' -ACCTCAGGGCCAAGGCTCCCGTGTGCCATTCTGC <b>g</b> T <b>g</b> tACCTCC-3'
16		3' -TGGAGTCCCGGTTCGAGGGGCACACGGTAAGAC <b>g</b> C <b>a</b> g <b>ca</b> tGGAGG-5'
17	Pe1	5' -TCTCCGAGCAATGGAGCCATTGTGGAGGGG-3'
18		3' -TCTGGCTCGTTACCTCGGTAACACCTCCCC-5'
19	Pe1M1*	5' -TCTCCGAG <b>gg</b> AT <b>c</b> GAGCCATTGTGGAGGGG-3'
20		3' -TCTGGCT <b>cc</b> T <b>a</b> g <b>t</b> CGGTAACACCTCCCC-5'
21	Pe1M2	5' -TCTCCGAGCAATGGAG <b>gg</b> T <b>t</b> cGGAGGGG-3'
22		3' -TCTGGCTCGTTACCT <b>g</b> cc <b>a</b> ag <b>c</b> CTCCCC-5'
23	Pe1M3	5' -TCTCCGAG <b>gg</b> AT <b>c</b> GAG <b>gg</b> T <b>t</b> cGGAGGGG-3'
24		3' -TCTGGCT <b>cc</b> T <b>a</b> g <b>t</b> CG <b>cc</b> A <b>a</b> g <b>c</b> CTCCCC-5'
25	Pe1M4*	5' -TCTCCGAGCAATGG <b>t</b> aCCATTGTGGAGGGG-3'
26		3' -TCTGGCTCGTTAC <b>ca</b> tGGTAACACCTCCCC-5'
27	HMG-cons	5' -ACACTGAGAACAAGCGCTCTCACAC-3'
28		3' -TGTGACTCTTGTTCGCGAGAGTGTG-5'
29	Sox9-cons	5' -GATCCGGACTAATAACAATAAAGTCGACG-3'
30		3' -CTAGGCCTGATTATTTGTTATTTTCAGCTGC-5'
31	Col48	5' -CTGTGAATCGGGCTCTGTATGCGCTTGAGAAAAGCCCCATTCATGAGA-3'
32		3' -GACACTGATCCCGAGACATACCGGA <b>ACTCTTTTCGGGGTAAGTACTCT</b> -5'

**Table 1. List of oligonucleotide primers.** \*-Symbolizes primers that were also used for site directed mutagenesis of luciferase constructs. **a, c, g, t**-small bold letters depict mutagenized nucleotides.

### Chemicals and enzymes

Commonly used chemicals were purchased from SIGMA, FLUKA, MERCK and SERVA companies.

Restriction enzymes used for subcloning and mapping were purchased from Fermentas and New

England Biolabs. T4 Polynucleotide Kinase (PNK), Calf Intestine Alkaline Phosphatase (CIAP), Klenow fragment, S1 nuclease, T4 polymerase, Tag polymerase, Pfu polymerase, and T7 Sequenase were purchased from Fermentas, Amersham and Promega.

### **Bacterial strains used for clonings or protein expression**

-DH5 $\alpha$ F- *recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac* [F $\phi$ proAB lacIqZDM15]

-Sure-e14-(McrA-)  $\Delta$ (mcrCB-hsdSMR-mrr)171*end A1sup E44thi-1gyr A96rel A1lacrec Brec Jsbc Cumu CTn5(Kan<sup>r</sup>) uvrC* [F' proAB lacI<sup>q</sup>Z $\Delta$ (M15 Tn10 (Tet<sup>r</sup>)]

-BL21-CodonPlus (DE3)-RIL:E.coliBF- *ompT hsdS(rB- mB-) dcm+ Tetr gal  $\lambda$ (DE3) endA Hte* [argU ileY leuW Camr]

### **Plasmid vectors**

- pBluescript II SK(+) or pBluescript II KS (+): (Stratagene), cloning vector.
- pCAT3<sup>®</sup>Control Vector: (Promega) used for transient cotransfection experiments.
- pTKCAT: (Promega) used for transient transfection experiments.
- pGL3-Basic: (Promega) used for cloning of regulatory regions and transient expression assays.
- pGL3-Control Vector: (Promega) used as well for transient expression assay.

### **Culture media for eukaryotic cells**

Culture media including DMEM (Dulbecco's Modified Eagle's Medium), Nutrient Mixture F-12 HAM, Fetal Calf Serum (FCS) and other reagents used for cell culture experiments were purchased from SIGMA-Aldrich Co and GIBCO BRL Life Technology.

### **Cell culture preparations**

Preparation of primary cultures of chicken embryo chondrocytes (CEC) and chicken embryo fibroblasts (CEF) from day 14 and day 8-10 embryos, respectively, has been described [138]. Briefly:

-Primary *chicken embryo chondrocyte (CEC)* cultures were obtained from sterna of day 14,5 embryos using 0.1% collagenase treatment, followed by plating the cell suspension at a density of  $7.5 \times 10^6 - 1 \times 10^7$  cells/60 mm plastic plates in DMEM supplemented with 5% FCS on the next day.

-Primary *chicken embryo fibroblast (CEF)* cultures were prepared from 8-10 days embryos. Following trypsin treatment fibroblasts were collected and on the next day, they were cultured in a suspension at a density of  $5 \times 10^6$  cells/60 mm plastic plates in DMEM containing 5% FCS.

-The mouse fibroblast cell line *NIH3T3* was obtained from American Type Culture Collection and maintained in DMEM containing 10% FCS. At confluency of  $5 \times 10^5$ , the cells were trypsinized and the proteins were extracted for the electrophoretic mobility shift assay (EMSA) experiment.

### **Sequence analysis**

Sequence analysis was made by E. Barta. All sequences were obtained either from the ENSEMBL genome browser [141], or from the EMBL databank [142]. Sequence manipulations were performed using the programs of the EMBOSS package [143]. Multiple alignments were made by the DIALIGN2 program [144] and were further improved by hand. Conserved motifs were searched with the MEME motif discovery program [145].

### **Isolation of genomic DNA from tissue cultures**

CEC and CEF cultures were subjected to the following procedure. The cells growing in monolayers were washed twice with ice-cold phosphate buffered saline (PBS: 10mM Na-phosphate and 0.15M NaCl, pH7.5). Then 2.5 ml buffer B (150mM NaCl, 5mM EDTA, pH7.8) and 2.5 ml buffer C (20mM Tris-Cl pH8, 20mM NaCl, 20mM EDTA, 1% SDS) containing 600 µg/ml proteinase K was added, followed by an incubation at 37°C for 1-2 hour. The viscous suspension was harvested, and twice phenol-chloroform extracted. Fifty µl of 10 mg/ml RNaseI (Sigma) was added to the sample, incubated at 37°C for 60 min, then extracted twice with phenol-chloroform, twice with chloroform and finally precipitated by ethanol.

### ***In vivo* footprinting with dimethyl sulfate and UV light**

*DMS, UV and Piperidine treatments:*  $7.5 \times 10^6$  -  $1 \times 10^7$  CEC and  $5 \times 10^6$  CEF cells were treated with a medium containing 0.2% dimethylsulfate (DMS) for 5 min at 22°C, or irradiated with 2400 J/m<sup>2</sup> UV light *in vivo*. The treatments were stopped by washing 2 times with 10 ml ice-cold PBS, and genomic DNA was isolated from the cultures treated *in vivo* as described earlier. As an *in vitro* control, genomic DNA was first isolated from untreated CEC and CEF. Fifty µg naked DNA was treated with 0.2% DMS in 200 µl of 50mM Na-cacodylate, 1mM EDTA, pH8 buffer for 5 min at 22°C. The reaction was stopped with 50 µl of 1.5M Na-acetate, pH7, 1M β-mercaptoethanol. As an *in vitro* control for UV treatment, genomic DNA was first extracted from untreated CEC or CEF cultures, dissolved at 0.5 µg/µl concentrations in TE. Ten µl drops on parafilm was irradiated with 4800 J/m<sup>2</sup> UV light. Thirty µg of DMS- and UV-modified *in vivo* and *in vitro* DNA samples were cleaved with 100 µl 1M piperidine for 30 min at 88°C, precipitated, washed twice with 80% cold ethanol, resuspended in 100 µl water and vacuum-dried in a Speed-vac concentrator (SVC-100H). The pellet was resuspended in 30 µl TE.

*Alkaline gel electrophoresis:* The size of DNA fragments generated by cleavage reactions was checked on 2% agarose gel made of 50mM NaCl, 1mM EDTA, and 2% agarose. Gel was presoaked overnight in the alkaline running buffer (50mM NaOH, 1mM EDTA). Two µl of each piperidine treated DNA was loaded on the gel and run with 50V. The gel was neutralized for 30 min in 1M Tris-HCl, pH7.6, 1.5M NaCl, stained for 30 min with 5 µg/ml ethidium bromide, destained with water, and photographed. The optimal size of the chemically cleaved DNA ranges between 200 bp to 1500 bp.

*Ligation mediated-PCR (LM-PCR):* The chemically cleaved DNA of proper size was amplified by LM-PCR [146] between positions -227 and +140 using gene-specific nested primers PU1 and PU2 for the upper strand; PL1 and PL2 for the lower strand; and linker primers LP11 and LP25 (Table 1). First a gene-specific primer (PU1 or PL1) annealed to the cleaved and denatured genomic DNA was extended with Sequenase<sup>TM</sup> and the ds linker was blunt end ligated. Then the fragments were PCR amplified using LP25 and PU2 or PL2. Then the PCR ladders were separated on sequencing gels with G+A, C+T sequence ladders, transferred to nylon membranes and hybridized with [<sup>32</sup>P]-labeled single-

stranded PCR probes made with gene-specific primers PU3 or PL3 (Table 1). G+A, C+T ladders were obtained by chemical cleavage according to the Maxam Gilbert procedure [147].

*Linker preparation:* 100 µl of 100 µM linker primers, LP11 and LP25 (Table 1), were annealed in 0.5 ml buffer containing 2M Tris-Cl (pH7.7) by incubating 3 min at 95°C, 1 min at 70°C and gradually cooling to room temperature, followed by over night incubation at 4°C. Annealed linkers were aliquoted and stored at -20°C until usage.

### **Preparation of cell extracts**

Crude cell extracts were prepared from CEC, CEF, and NIH3T3 cultures for electromobility shift assay (EMSA). Cells collected and washed twice in ice-cold PBS, were lysed in 20mM Hepes, pH7.9, 350mM NaCl, 0.5mM EDTA, 0.5mM EGTA, 2mM dithiothreitol (DTT), 0.2mM Phenylmethylsulphonylfluoride (PMSF) and 1% Nonidet P-40 on ice for 30 min. The chromatin was centrifuged for 5 min at 20,800xg at 4°C. The supernatants were supplemented with 10% glycerol, aliquoted and stored at -80°C.

### **Synthesis and purification of bacterially expressed GST-SOX9**

The GST-SOX9 vector for bacterial expression of the fusion of glutathione S-transferase (GST) with full length SOX9 and the empty control vector were kindly provided by P. Berta [124]. The recombinant protein was produced in bacterial strain BL21-CodonPlus (DE3)-RIL after induction with 0.1mM isopropyl-β-D-thiogalactopyranoside (IPTG) for 2 h at 37°C. The cells were harvested, resuspended in 150mM NaCl, 1mM DTT, 5mM EDTA, 25% sucrose, and 50mM Tris, pH7.5 and sonicated for 2 min at 4°C. After centrifugation, the cleared lysate was applied onto glutathione-Sepharose<sup>TM</sup> 4B (Amersham Biosciences, Upsala, Sweden) beads and washed three times in 250mM NaCl, 5mM EDTA, 50mM Tris, pH7.6 and three times in 120mM NaCl, 5mM EDTA, 50mM Tris, pH7.6. Purified proteins were eluted from the matrix with the latter buffer supplemented with 10mM reduced glutathione by incubating for 30 min at 4°C. Yields were tested by SDS-PAGE.

### **EMSA and supershift experiments**



Crude cell extracts of CEC, CEF and NIH3T3 or bacterially synthesized SOX9 were incubated in the presence of 100 ng of poly(dG-dC)poly(dG-dC) and various cold competitors with 10-20 fmol of [ $\gamma$ - $^{32}$ P] ATP end-labeled DNA probe, and loaded on prerun 5-6.6% SDS-polyacrylamid gel (SDS-PAGE) in 0.25xTBE buffer as described [139]. Supershift experiments were performed as described [139] with antisera raised against SOX9, L-Sox5 and Sox6 [29] kindly provided by B. de Crombrughe and V. Lefebvre.

### **Luciferase reporter gene constructs for transient expression assay**

The *Ava*I (blunted)-*Nco*I fragment harboring the matrilin-1 minimal promoter between positions -15 to +67 was inserted into the *Sma*I and *Nco*I site of the pGL3-Basic vector (Promega) to produce *MpLuc*. In construct *4xPe1-MpLuc*, four copies of the Pe1 element (Table 1) subcloned in the *Eco*RV site of pBluescript II SK(+) (Stratagene) was excised with *Sac*I-HindIII (blunted) and ligated to the *Sac*I-NheI (blunted) site of *MpLuc*. The *Sac*I-*Nco*I fragment including the proximal promoter region of chicken matrilin-1 gene between -334 to +67 was ligated to the *Sac*I and *Nco*I site of the pGL3-Basic vector to produce the *FO15Luc* construct. *FO15Luc* derivatives  $\Delta$ *Pe1M1*,  $\Delta$ *Pe1M4*,  $\Delta$ *Ine1M1*,  $\Delta$ *Ine1M2*, and  $\Delta$ *Ine1M3* were made by the PCR based QuickChange<sup>TM</sup> Site-Directed Mutagenesis (Stratagene) according to the manufacturer's instructions using *FO15Luc* as a template and oligonucleotides Pe1M1, Pe1M4, Ine1M1, Ine1M2 and Ine1M3 (Table 1) carrying point mutations in the Sox recognition sites and in the spacer region of Pe1, respectively. The *Bgl*III-*Nco*I matrilin-1 promoter fragment between -1137 and +67 was inserted into the *Bgl*III and *Nco*I sites of pGL3-Basic to obtain pCMPLuc. The structure of all the constructs and the accuracy of the mutagenesis were verified by restriction mapping and nucleotide sequencing.

### **Site-directed mutagenesis**

Two synthetic oligonucleotide primers containing the desired mutation, each complementary to opposite strands of the vector, were extended during temperature cycling by *Pyrococcus furiosus* (Pfu) DNA polymerase (Fermentas). After the cycling, the products were treated with DpnI restriction digestion to eliminate the methylated parental template. The nicked DNAs were then transformed into

DH5 $\alpha$ F *Escherichia coli* (*E. coli*) and repaired. Accuracy of the mutagenesis was checked by sequencing.

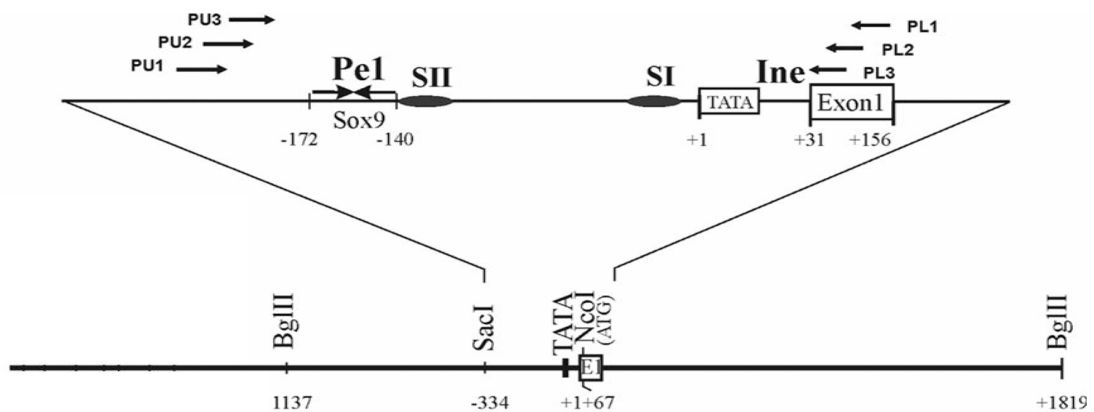
### **Transient expression assay**

CEC, CEF and HDM cultures were released by enzymatic digestion, seeded at a density of  $7.5 \times 10^6 - 1 \times 10^7$  cells/60 mm dishes 24 h prior to transfection. Cells were transfected with the calcium phosphate coprecipitation method as described [139] with 5  $\mu$ g of each luciferase reporter gene constructs. To test for transfection efficiency, 0.5  $\mu$ g of pCAT<sup>R3</sup> Control Vector (Promega) was included as an internal control, and an empty vector was added to have a total DNA amount of 10  $\mu$ g. Parallel plates were also transfected with pGL3-Control Vector (Promega). The transfected cells were harvested 48 h later by centrifugation and the cell pellet was lysed by two cycles of freezing and thawing in 200  $\mu$ l 50mM Tris-HCl, 2mM DTT, 0.5mM EDTA, 0.5mM EGTA, 1% NP-40. Luciferase activity was measured from the supernatant in a Luminoscan Ascent (ThermoLabsystem 2.6) using luciferin substrate (Promega) [148]. Relative luciferase activities were expressed in percent of that of *FOI5Luc* taken as 100%. All transfections were performed in duplicates with at least two different DNA preparations. Data are presented as means  $\pm$  S.E.M

## RESULTS

### Sequence analysis reveals several Sox-binding sites within the short promoter of the chicken *matrilin-1* gene

As the important role of Sox9, L-Sox5 and Sox6 has been demonstrated in the control of chondrogenic differentiation and the activation of cartilage-specific genes, we aimed to identify potential Sox-binding sites in the promoter control region of the chicken *matrilin-1* gene. The location and the sequence analysis of the putative Sox motifs are shown in Figure 6 and 7. Throughout 1.2 kb promoter region tested between positions -1137 and +64, we identified two potential areas, herein referred to a promoter element 1 (Pe1) and initiator element (Ine), that seem to contain several Sox-binding sites. The Pe1 element is located ~140 bp upstream of TATA and it includes a pair of inverted motifs highly similar to the AGAACAATGG motif, which was shown to be the preferred binding site of Sox9 *in vitro* [149] (Figure 7). The inverted Sox motifs are separated by 2 nucleotides within Pe1.



**Figure 6. Diagram of the promoter region of the chicken *matrilin-1* gene analyzed using *in vivo* and *in vitro* techniques.** The schematic figure of the 5' end of the gene depicts the location of conserved sequence blocks

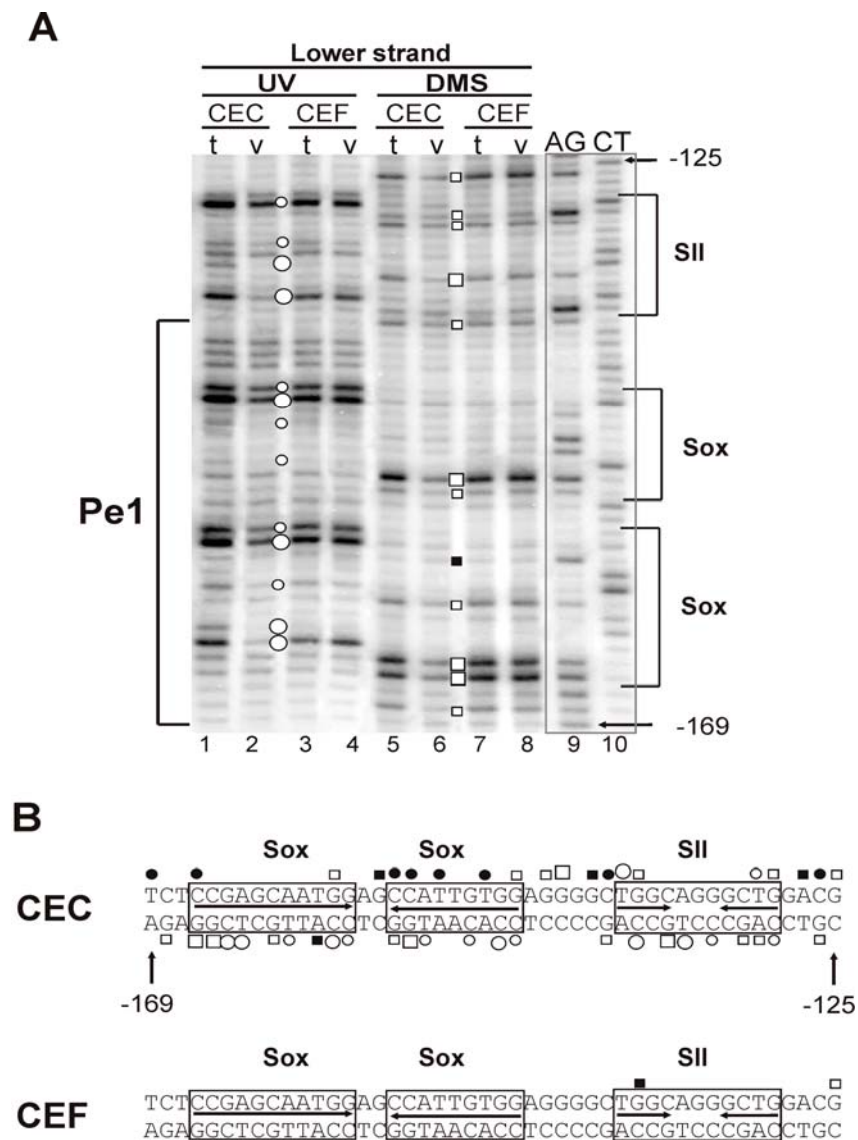


DNA was treated with DMS or UV light to modify G residues at N-7 position or produce (6-4) photoproducts at TC and CC dinucleotides, respectively. These modifications are sensitive to bound proteins, therefore the areas of protein-DNA interactions appear as footprints on LM-PCR genomic sequencing ladders [146]. LM-PCR analysis of naked CEC and CEF DNA treated with the same reagents *in vitro* served as a reference. Differences in the modification patterns between the *in vivo* and *in vitro* treated samples, apparent as hyporeactivities (protections) or hyperreactivities, indicated *in vivo* DNA-protein contacts at specific sequences. Sets of independent experiments revealed cartilage-specific binding of transcription factors to the short promoter (Figure 7 and 8). We focused on the vicinity of the Pe1 element between -169 and -125 (Figure 8), and Ine element between -30 and +52 (Figure 9) that harbors TATA box and the first transcription start site of the chicken gene.

Apart from Pe1, the first region also covered the SII element identified previously [139]. In chondrocytes, occupancy was observed at the NFI contact points of the SII element (Figure 7A, lanes 1, 2, 5 and 6). The conserved Pe1 between -169 and -139 was also clearly protected by bound transcription factors at certain G, CC and CT nucleotides of the inverted Sox motifs, while hypersensitivity was seen on the opposite strand. Results of DMS and UV footprinting obtained on the upper and lower strands of DNA (Figure 7A and data not shown) are summarized in Figure 8B. Furthermore, hyporeactivity indicated factor binding at G residues of an Sp1-like motif between the Pe1 and SII elements (Figure 8) and also at G residues of several Sp1-like elements located between SII and the TATA motif (data not shown). These data demonstrate *in vivo* occupancy of potential recognition sequences for Sox and certain ubiquitous factors as well within Pe1 and its vicinity in chondrocytes. As opposed to this, the complete absence of footprints in the short promoter in fibroblasts in repeated experiments indicated no factors bound to their recognition sequences in the non-expressing cell type (Figure 8A, lanes 3, 4, 7 and 8, Figure 8B).

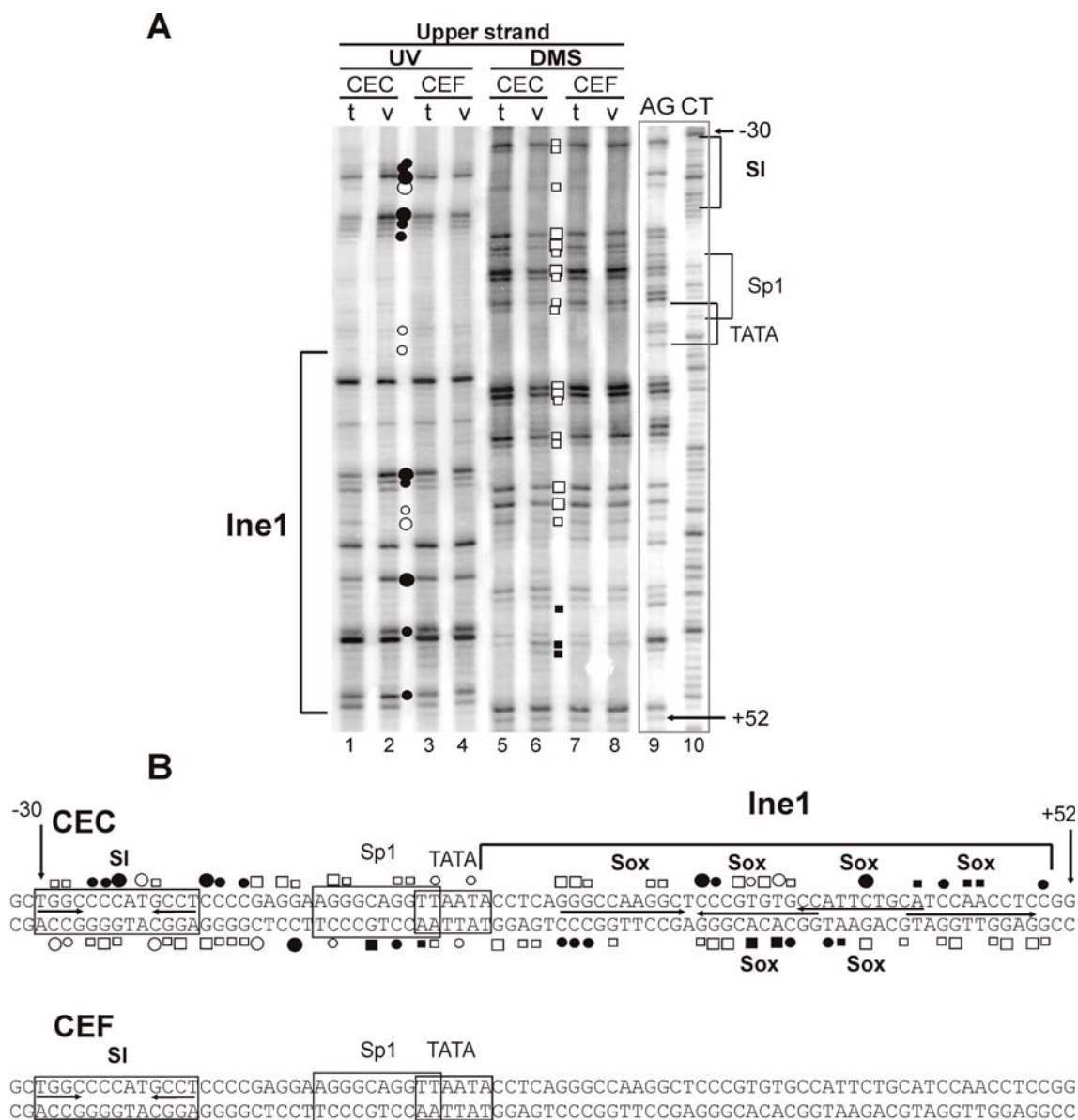
Similar results were obtained in the other region between -30 and +52 harboring the vicinity of the TATA box (Figure 9). This area included the SI element which was identified previously [139], the putative Sp1 site, the TATA box, the conserved motif at the first transcription initiation site, and number of potential Sox CA/TTTGA/TA/T motifs [149] downstream to TATA. Inspection of the sequence revealed two pairs of inverted putative Sox motifs with 5/10, 6/10, 7/10 and 4/9 identity with

the consensus Sox-binding sites. Following treatments by either UV light or DMS, CECs exhibited a number of CC and CT- or A- and G-residues of much higher or lower relative intensity, whereas no footprints were observed for CEFs. Similar results were obtained for the other strand (data not shown), and all data obtained are summarized in Figure 9B.



**Figure 8. Cartilage-specific occupancy of the Pe1 and SII elements in genomic footprinting** A) Footprints are shown between positions -169 and -125 on the lower strand of the DNA. AG and CT are Maxam-Gilbert control sequences. DNA from CEC and CEF cultures treated *in vivo* (v) with DMS (open and closed boxes) or UV light (open and closed circles) is compared with the *in vitro* (t) DNA samples treated with these reagents after isolation from CEC and CEF. Differences in the modification patterns between 'v' and 't' treatments,

visible as hyperreactivities (large and small closed circles or closed boxes) or protections (large and small open circles or open boxes), indicate *in vivo* DNA-protein contacts at specific sequences. **B)** Summary of the *in vivo* footprinting data is shown on both strands. The previously identified NFI-binding site in the SII element [139] and the inverted repeat harboring the putative paired Sox-binding sites are boxed.



**Figure 9. Cartilage-specific occupancy of the Ine1 and SI elements in genomic footprinting** **A)** Footprints are shown between positions -30 and +52 on the upper strand of the DNA. AG and CT are Maxam-Gilbert control sequences. DNA from CEC and CEF cultures treated *in vivo* (v) with DMS (open and closed boxes) or UV light (open and closed circles) is compared with the *in vitro* (t) DNA samples treated with these reagents after isolation from CEC and CEF. Differences in the modification patterns between 'v' and 't' treatments,

visible as hyperreactivities (large and small closed circles or closed boxes) or protections (large and small open circles or open boxes), indicate *in vivo* DNA-protein contacts at specific sequences. **B)** Summary of the *in vivo* footprinting data is shown on both strands. The previously identified NFI-binding site in the SI element [139], putative Sp1 site and the TATA motifs are boxed. Potential Sox-binding sites are shown by arrows under Ine1 label.

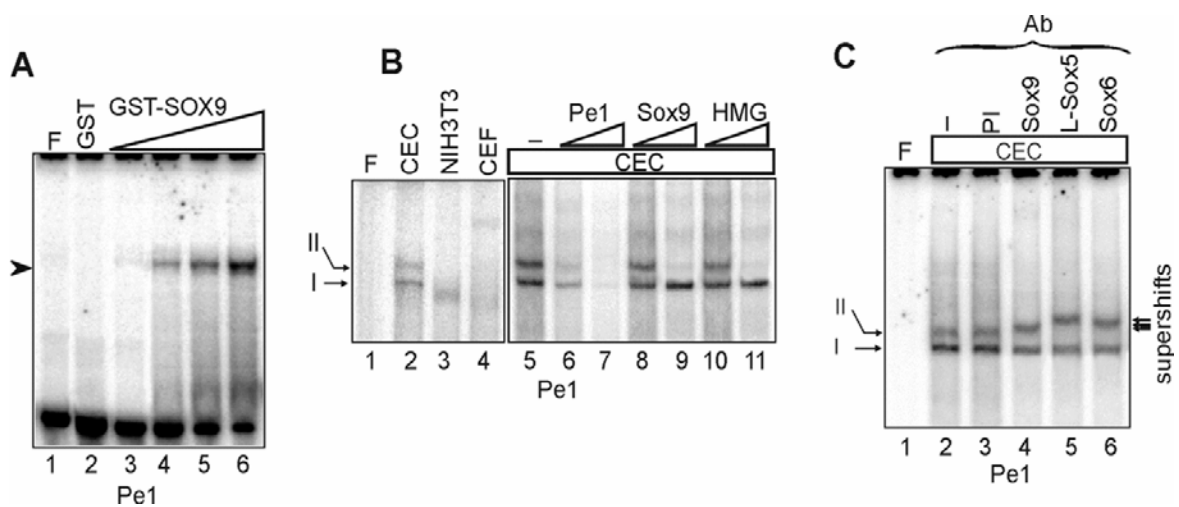
To sum up, our data indicate *in vivo* cartilage-specific occupancy at potential recognition sequences for Sox factors within Pe1 and Ine1, as well as for ubiquitous factors within the entire short promoter in chondrocytes. The complete absence of footprints in the short promoter in fibroblasts indicates that neither cartilage-specific, nor ubiquitous factors bound to their recognition sequences in the non-expressing cell type. This suggests that regulation at the chromatin level can be involved in the activation of the gene in chondrocytes.

### **Sox9, L-Sox5 and Sox6 interact with the Pe1 element *in vitro***

The Sry-type, high-mobility group (HMG)-box containing transcription factor SOX9 serves as a master regulator of the chondrocyte lineage. Therefore, we aimed to identify potential Sox-binding sites in the regulatory region of the matrilin-1 gene. The inverted repeats within the conserved Pe1 element, which exhibited cartilage-specific protection in genomic footprinting, share 7/10 bp identity with the consensus Sox9-binding site (Figure. 7). To confirm that Sox9 can indeed bind to Pe1 in DNA-binding assay *in vitro*, oligonucleotides corresponding to this region (between –169 and –139) were synthesized for both strands, and bacterially expressed and purified GST-SOX9 fusion protein was incubated with the element. As demonstrated in Figure 10A, the recombinant SOX9 formed a nucleoprotein complex efficiently on the element (lanes 3-6). When EMSA was performed with nuclear proteins from CEC, NIH3T3 and CEF cultures, we noticed tissue-specific differences in the pattern and behavior of complexes (Figure 10B, lanes 1-4). Two specific DNA-protein complexes absent from fibroblasts were formed with CEC nuclear proteins (lane 2). Competition EMSA revealed that ds oligonucleotides harboring consensus Sox or HMG box motifs interfered only with the formation of the slowly migrating complex II (lanes 8-11). The same complex was supershifted with antibodies specifically recognizing Sox9, L-Sox5 and Sox6 (Figure 10C), indicating that each of the



cartilage-specific Sox transcription factors participated in the formation of complex II on Pe1 in chondrocytes. On the other hand, the observation that complex I was neither competed nor supershifted with Sox-specific oligonucleotides or antibodies, respectively, indicated that the element also interacted with other transcription factors in the expressing cell type. Together, these data provide sufficient evidence to conclude that the chondrocyte-specific *in vivo* footprints at the inverted Sox motifs of Pe1 might be due to the binding of Sox9, L-Sox5 and Sox6 proteins. In addition, the Pe1 element can also form a complex with an unrelated chondrocyte nuclear protein.

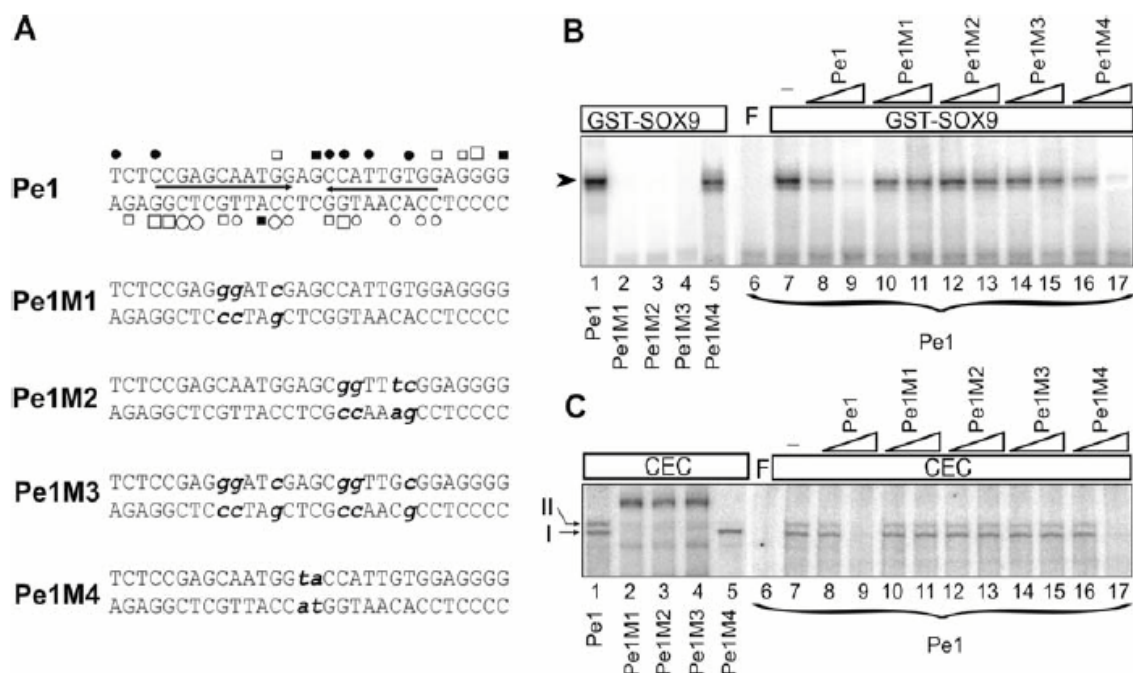


**Figure 10. Analysis of the interaction of the Pe1 element with Sox proteins *in vitro*** A) [<sup>32</sup>P]-labeled Pe1 was incubated with 2.0 μg of GST alone (lane 2) and increasing amounts (0.25, 0.5, 1.0 and 2.0 μg) of purified GST-SOX9 (lanes 3-6). No protein was added to lane F (free probe). B) Radiolabeled Pe1 was incubated with 3 μg of nuclear proteins extracted from CEC, CEF and NIH3T3 cells (lanes 2-4). The binding of chondrocyte nuclear proteins was competed with 50- and 500-fold molar excesses of Pe1, consensus Sox9 and HMG elements (lanes 6-11). No competitor was added to lane 5. C) Supershift experiment was performed without antiserum (lane 2) and with 1 μl each of preimmune antiserum (PI) (lane 3) and antibodies specifically recognizing Sox9, L-Sox5 and Sox6 (lanes 4-6). The arrows point to the supershifted complexes.

### Mutations in either of the paired Sox motifs of Pe1 interfere with protein binding in EMSA

To examine further the Pe1 element, point mutations were introduced into either one (Pe1M1 and Pe1M2) or both (Pe1M3) of the inverted nonameric Sox motifs, or the spacer region between them (Pe1M4) (Figure 11A). These oligonucleotides were used as probes in EMSA with purified

recombinant SOX9 and CEC nuclear proteins. Mutants Pe1M1, Pe1M2 and Pe1M3 did not bind to GST-SOX9 (Figure 11B, lanes 1-4) and did not compete for binding of Pe1 to either purified recombinant SOX9 or CEC nuclear proteins (Figures 11B and 10C, lanes 6-15). In other words, point mutations in either of the paired Sox motifs prevented the interaction between GST-SOX9 and Pe1, indicating that both nonameric Sox motifs within the pair were essential for recognition by SOX9. Consistent with earlier observations [150], these results clearly demonstrate the importance of paired Sox sites in cartilage-specific gene regulation. Mutations M1, M2 and M3 also equally interfered with the formation of CEC nucleoprotein complexes I and II, but yielded a very slowly-migrating complex (Figure 11C, lanes 2-4).

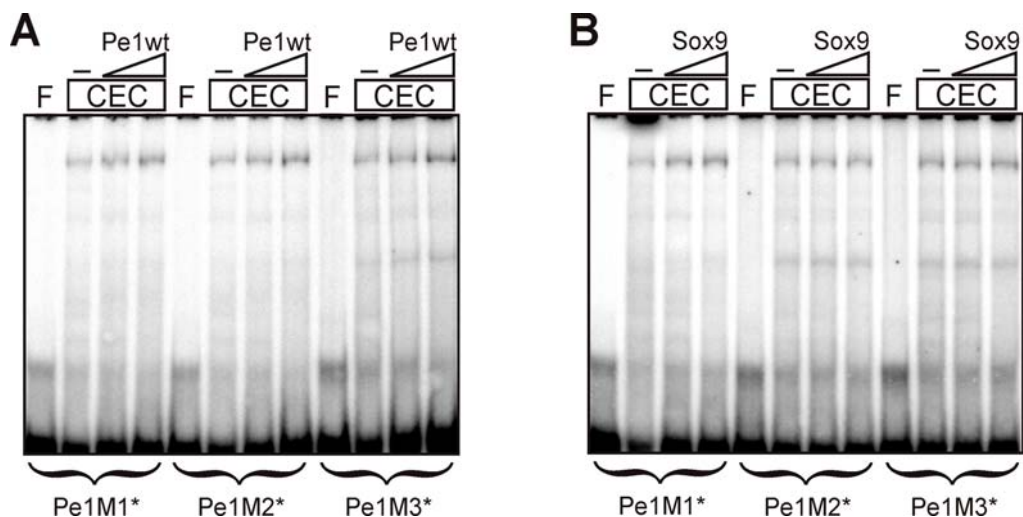


**Figure 11. Mutational analysis of Pe1 by EMSA** A) Nucleotide sequences of the wild type and mutant versions of Pe1. The inverted repeats (arrows) harboring the paired Sox motifs as well as protections and hiperreactivities are indicated as in Figure 5. Point mutations presented in bold lower case letters were introduced at nucleotides, which showed *in vivo* occupancy at the Sox motifs or in the spacer region. B) Interaction of the wild type (lane 1) and mutant versions (lanes 2-5) of radiolabeled Pe1 with 2.0  $\mu$ g of purified GST-SOX9. Binding of radiolabeled Pe1 to GST-SOX9 was competed by 50- and 500-fold molar excess of cold normal and mutant Pe1 as indicated on the top (lanes 8-17). C) EMSA was performed to compare the CEC nucleoprotein complexes formed on the wild type and mutant Pe1 elements (lanes 1-5). Formation of CEC nucleoprotein

complexes was competed with 50- and 500-fold molar excess of cold Pe1 and its mutant versions as indicated on the top (lanes 8-17). No competitor was added to lane 7. F, free probe.

Since competition with Sox-binding site or wild-type Pe1 element did not diminish the latter complex (Figure. 12A and 12B), we concluded that the mutagenesis possibly created a binding site for an unknown nuclear factor synthesized in chondrocytes. However, the unknown nuclear factor did not recognize the wild-type element. The observation that mutations in either one of the inverted Sox motifs disrupted not only the Sox-specific complexes II, but complex I as well suggests that the formation of the two complexes is not completely independent. This gives rise to two possibilities. One explanation can be that the mutations also disrupted the overlapping binding site for a currently unidentified factor forming complex I. The alternative, and more likely explanation can be that binding of Sox9 homodimers to the paired sites can cause bending of the DNA, and thereby may promote an otherwise very weak interaction between Pe1 and the unidentified factor. Thus, abolishment of Sox9-binding to Pe1 may prevent the binding of the unidentified factor as well.

Pe1M4 was able to compete with the wild-type element for binding to recombinant SOX9 or CEC nuclear proteins (Figures 11B and 11C, lanes 16-17). However, the spacer mutation reduced the binding efficiency of Pe1M4 to GST-SOX9 (Figure 11B, lane 5). Contrary to Pe1, Pe1M4 formed only a single complex of altered mobility with CEC nuclear proteins in repeated experiments (Figure 11C, lane 5). This indicates that although the spacer mutation did not abolish the interaction of Pe1 with GST-SOX9, it modified the formation of CEC multiprotein complexes on the element. The dramatic effect of point mutations supports the conclusion that Pe1 can interact with Sox9 *in vitro* at the nonameric palindrome and suggests the binding of an unknown factor to the spacer region.



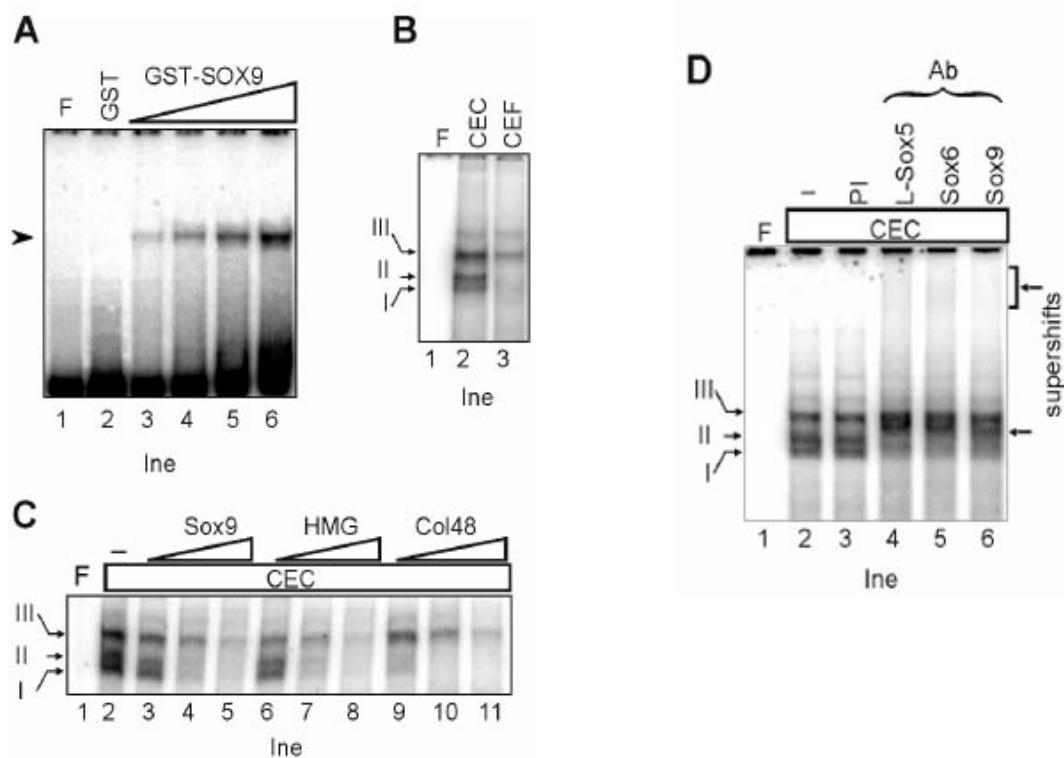
**Figure 12. Pe1 mutants in competition EMSA with A) wild type Pe1, and B) consensus-Sox9.** Cold ds Pe1 or consensus-Sox9 were added in 50- and 500-fold molar excesses as indicated on the top in to the Pe1M1/CEC to Pe1M3/CEC nucleoprotein complexes. No competitor was added to lane ‘-’. F, free probe.

**The initiator element Ine can also interact with Sox factors *in vitro***

As genomic footprinting indicated high occupancy of Sox motifs in the vicinity of the transcription start sites of the gene *in vivo* (Figure 9B), we addressed the question, whether nuclear factors could bind to this area *in vitro*. To answer this question, first we performed EMSA on the initiator element (Ine) between positions +6 to +50 with purified recombinant SOX9 and nuclear proteins extracted from CEC and CEF cultures (Figure 13A and B). Specific binding of GST-SOX9 but not GST alone to Ine demonstrated that Sox factors could indeed interact with the element *in vitro* (Figure 13A). With chondrocyte nuclear proteins, three major nucleoprotein complexes were formed. Complexes I and II were absent, however, when CEF nuclear extracts were used, indicating that those were formed by cartilage-specific transcription factors (Figure 13B). The latter complexes were displaced with HMG- or Sox9-specific competitors, as well as with the 48-bp *Col2a1* enhancer element containing Sox9 and L-Sox5/6 recognition sequences, indicating that Sox factors may participate in their formation (Figure 13C). To test, whether the chondrogenic transcription factors Sox9, L-Sox5 and Sox6 were involved in the complex formation, we performed supershift experiments using antibodies specifically recognizing Sox9, L-Sox5 and Sox6. Complexes I and II were supershifted providing direct evidence that all three chondrogenic Sox factor can bind to the element *in vitro* (Figure 13D). Complex III, however, was neither efficiently displaced with

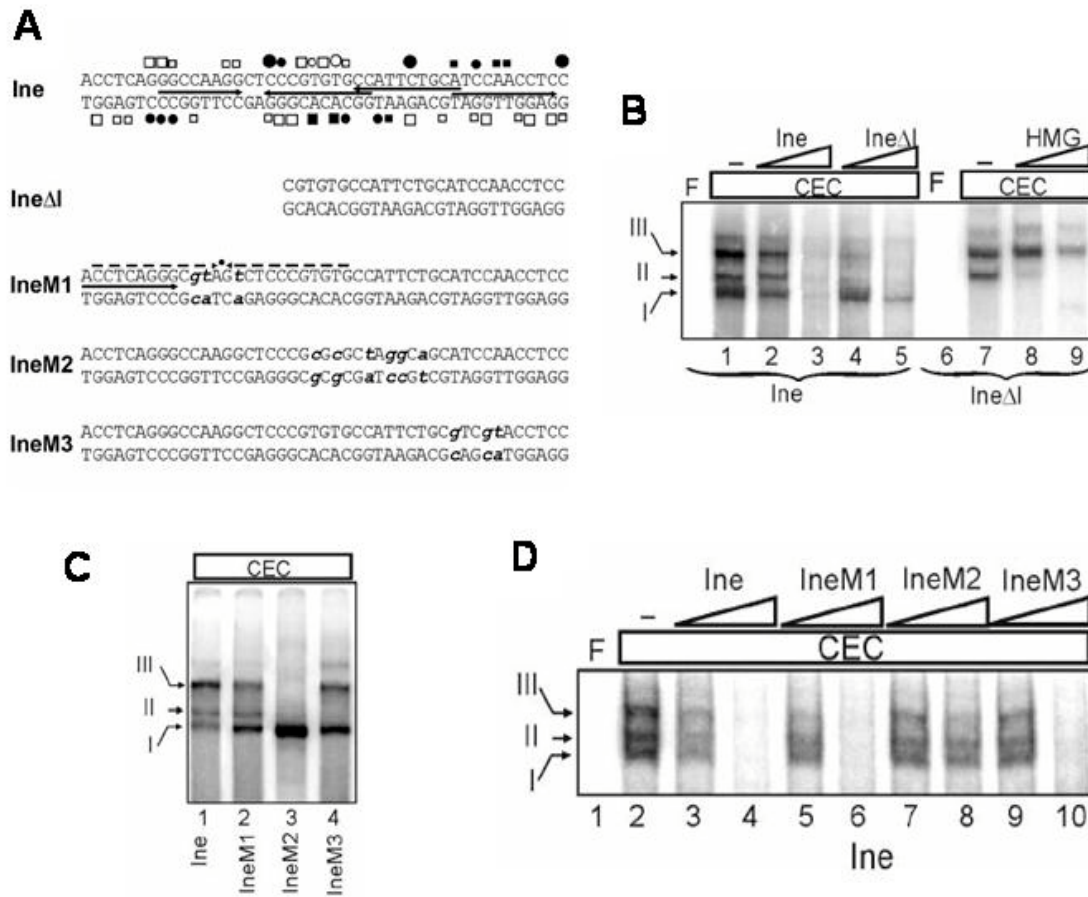
competitors harboring Sox recognition sequences, nor supershifted with Sox-specific antibodies (Figure 13C and 13D). A complex of the same mobility was also present in the non-expressing CEF samples (Figure 13B), suggesting that complex III may contain a ubiquitous factor present in many cell types.

To confirm that mutation of the Sox motifs interfere with complex formation, we performed EMSA with mutant versions of Ine (Figure 14A). When a shorter version of the initiator element Ine $\Delta$ I lacking the first pair of Sox motifs, but harboring the consensus initiator motif was incubated with chondrocyte nuclear proteins, only two major complexes appeared with apparent mobilities resembling to those of complexes II and III (Figure 14B, lane 7). Based on competition EMSA, only the former, faster migrating complex was efficiently displaced with HMG-specific oligonucleotides (Figure 14B, lanes 7-9). In consistence with this observation, Ine $\Delta$ I was also able to efficiently interfere only with the formation of complexes II and III on Ine1 (Figure 14B, lanes 4,5). The lack of abolishment of the complex I by Ine $\Delta$ I competitor indicated that this complex was formed on the first pair of Sox motifs of Ine.



**Figure 13. Analysis of the interaction of the Pe1 element with Sox proteins *in vitro*** **A)** [<sup>32</sup>P]-labeled Pe1 was incubated with 2.0 μg of GST alone (lane 2) and increasing amounts (0.25, 0.5, 1.0 and 2.0 μg) of purified GST-SOX9 (lanes 3-6). No protein was added to lane F (free probe). **B)** Radiolabeled Ine was incubated with 3 μg of nuclear proteins extracted from CEC and CEF cells (lanes 2,3). **C)** The binding of chondrocyte nuclear proteins to the Ine was competed with 50- and 500-fold molar excesses of consensus Sox9, HMG, and Col48 that contain recognition sequences of L-Sox5/6/and Sox9. No competitor was added to lane 2. **D)** Supershift experiment was performed without antiserum (lane 2) and with 1 μl each of preimmune antiserum (PI) (lane 3) and antibodies specifically recognizing Sox9, L-Sox5 and Sox6 (lanes 4-6). The arrows point to the supershifted complexes.

In agreement with this conclusion, IneM1 carrying point mutation in a second motif of the first inverted Sox pair that is missing from IneΔI, affected the formation of complex I, by increasing the intensity of this complex (Figure 14C, lane 2). This mutation increased the length of the palindrome and thus the binding efficiency of Sox factors, as Sox proteins are known to recognize four-way junction sequences [151]. IneM1 mutation did not decrease the ability of IneM1 competitor to efficiently displace both the Sox-specific and unrelated complexes (Figure 14D, lanes 5,6). Moreover, this mutation decreased the intensity of complex III (Figure 14C, lane 2). Mutation in IneM2 disrupted both pairs of inverted Sox motifs. As a consequence only a single complex appeared with mobility slightly below that of complex I (Figure 14C, lane 3), but it was not related to that, because IneM2 did not displace either of the complexes formed on Ine (Figure 14D, lanes 7,8).



**Figure 14. Mutational analysis of Pe1 by EMSA** **A)** Nucleotide sequences of the wild type and mutant versions of Ine. The inverted repeats (arrows) harboring the paired Sox motifs as well as protections and hyperreactivities are indicated as in Figure 5. Point mutations presented in bold lower case letters were introduced at nucleotides, which showed *in vivo* occupancy at the Sox motifs. **B)** Competition EMSA was performed to compare the CEC nucleoprotein complexes formed on the wild type Ine (lanes 1-5) and IneΔI (lanes 7-9) elements. Formation of Ine/CEC or IneΔI/CEC nucleoprotein complexes were competed with 50- and 500-fold molar excess of cold Ine (lanes 2,3) and cold IneΔI (lanes 4,5) or cold HMG (lanes 8,9) as indicated on the top. No competitor was added to lane 1 and 7. F, free probe. **C)** Interaction of the wild type (lane 1) and mutant versions (lanes 2-5) of radiolabeled Ine with 3 μg of nuclear proteins extracted from CEC. **D)** Formation of Ine/CEC nucleoprotein complexes was competed with 50- and 500-fold molar excess of cold Ine and its mutant versions as indicated on the top (lanes 3-10). No competitor was added to lane 2. F, free probe.

The observation that complex III did not form on IneM2 indicates that the mutation also disrupted the recognition site for the unidentified non-cartilage-specific protein. Alternatively, the binding of this factor may be dependent on the binding of Sox proteins to Ine and bending the DNA. Thus, the disruption of both pairs of inverted Sox motifs may prevent the Sox-facilitated binding of this unknown factor as well.

On the other hand IneM3 carrying mutation in the second inverted repeat did not produce the Sox-specific complex II (Figure 14C, lane 4), but it competed for the formation of Sox-specific and unrelated complexes as well (Figure 14D, lanes 9,10). Disruption of complex II in IneM3 increased the intensity of complex I (Figure 14C, lane 4). As mutations in the putative Sox motifs interfered with the formation of complexes, this further supported the recognition of Ine with Sox factors *in vitro*. Together, these data give sufficient evidence to conclude that, in addition to the tissue-specific binding ability, Ine contains Sox recognition sites, which can interact with Sox9, L-Sox5 and Sox6. We also conclude from the data that the chondrocyte-specific DMS and UV footprints observed at the inverted Sox motifs of the initiator region *in vivo* might be due to the binding of Sox9, L-Sox5 and Sox6 proteins harboring HMG-box DNA-binding domains.

### **Functional analysis of Pe1 and Ine elements in transient expression studies**

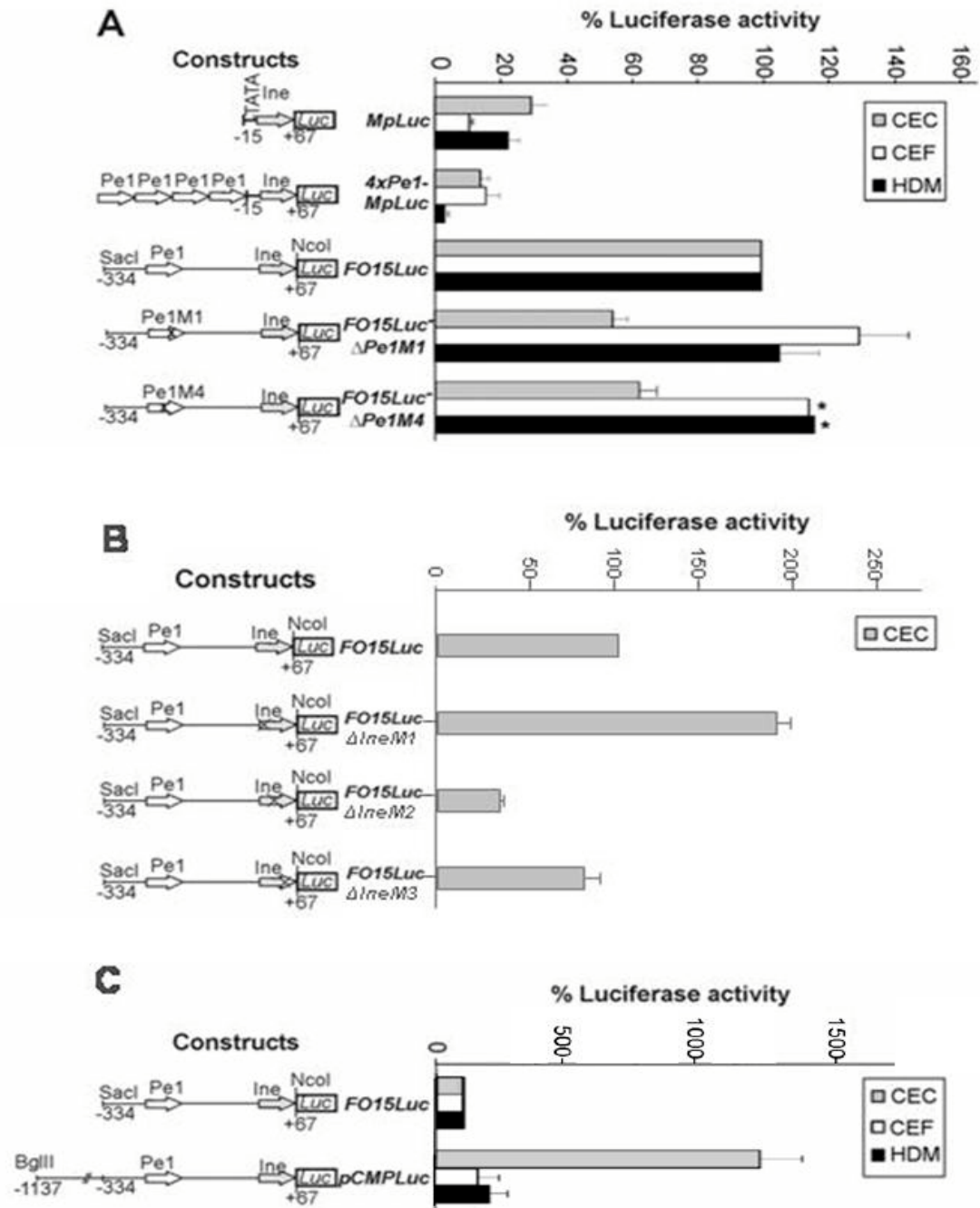
In order to determine which binding sites for Sox proteins are important for transcriptional activity of the chicken matrilin-1 gene, wild type and mutant versions of the promoter between positions -334 and +67 were fused to the luciferase reporter gene, and the promoter activity was measured in transient expression assays (Figure 15). In agreement with the data in transgenic mice [140], the minimal promoter between -15 and +67 (*MpLuc*) had a very low activity in expressing and nonexpressing cell types. Extension of the sequence up to -334 in 5' direction in construct *FO15Luc* elevated the activity of the short promoter 3.4-fold in chondrocytes (Figure 15A). Activation was also observed in CEF and HDM cultures, possibly due to the binding of ubiquitous factors (Sp1, NFI, etc) to the putative motifs of the short promoter in transient expression assays. Introducing mutations M1 into one of the paired Sox9 motifs of the Pe1 element by *in situ* mutagenesis in  $\Delta Pe1M1$  decreased the luciferase activity by almost 2-fold in chondrocytes (Figure 15A). The same mutation did not decrease or rather slightly increase the promoter activity in CEF and HDM cultures, thus demonstrating the tissue- and developmental stage-specific effect of mutation in the Sox-binding sites of Pe1. Interestingly, Pe1M4 containing mutations in the spacer region between the intact Sox motifs decreased the short promoter activity in derivative  $\Delta Pe1M4$  to 60 % in chondrocytes, but not in CEF and HDM cultures (Figure 15A). In agreement with the altered complex formation with chondrocyte



nuclear proteins in EMSA (Figure 11C), this suggests that Pe1 probably interacts with another protein in chondrocytes that is also critical for the activity of the element. Based on the functional assays, we conclude that Pe1 is likely to contribute significantly to the short promoter activity by binding to cartilage-specific Sox proteins via the inverted motifs and to a currently unidentified factor in the spacer region.

The Pe1 element is found within 140-170 bp upstream of the TATA motif in chicken gene (Figure 7) and harbors inverted nonameric Sox-binding sites that interact with Sox9, L-Sox5 and Sox6 *in vitro* (Figure 10). To test whether the element has an enhancer activity and/or the location is important for the function, four copies of the element were placed upstream of the minimal promoter in construct *4xPe1-MpLuc* (Figure 15A). However, this further decreased (CEC and HDM) or did not alter significantly (CEF) the luciferase activity as compared to the minimal promoter construct *MpLuc* (Figure 15A), suggesting that the Pe1 element is not an enhancer-like element and it cannot drive the promoter in such a close proximity to the TATA motif or it may work cooperatively with adjacent sequences binding to other factors.

To determine whether the Sox protein binding sites identified by EMSA within Ine are functionally relevant, we also tested the effect of Ine point mutations on the short promoter activity in transient expression assays in CEC (Figure 15B). The derivative of *FO15Luc* referred to as *ΔIneM1*, which carries point mutation M1 within the Ine element, exhibited almost 2-fold higher luciferase activity as compared to *FO15Luc*. This is in agreement with the increased formation of CEC nucleoprotein complex I *in vitro* binding assays (Figure 14C).



**Figure 15. Functional analysis of Pe1 and Ine in transient expression assays** Luciferase reporter constructs harboring the minimal or the proximal promoter region of the matrilin-1 gene (A&B) or promoter fragments of increasing length (C) between positions indicated are shown on the left. Construct *4xPe1-MpLuc* carries four tandem copies of Pe1 upstream of the minimal promoter. Point mutations of Pe1M1, Pe1M4, IneM1-M3, were introduced into *FO15Luc* in derivatives *Pe1M1*, and *Pe1M4* in (A) or *IneM1*, *IneM2* and *IneM3* in (B) respectively. Luciferase activities are expressed in % of that of the *FO15Luc* set at 100%. Values represent the averages  $\pm$  S.E.M. of 5-10 independent transfection experiments in CEC, CEF and HDM cultures. Asterisk mark the values shown as averages of three independent experiments.

On the other hand,  $\Delta IneM2$  had a 3-fold decrease in luciferase activity (Figure 15B), supporting that the disruption of both inverted Sox motifs within the initiator element not only abolished the interaction of Ine with Sox proteins in EMSA (Figure 14C), but indeed significantly decreased the short promoter activity as well. Finally, derivative  $\Delta IneM3$  harboring a modification in the second pair of inverted Sox motifs, which however did interfere with one of the Sox-specific nucleoprotein complex II *in vitro* (Figure 14C), exhibited 10-15% decrease in luciferase activity (Figure 15B).

When longer promoter fragment up to -1137 in 5' direction in construct *pCMPLuc* was tested in transient expression assays, we observed 12.4-fold enhancement of the reporter gene activity in CEC culture (Figure 15C). In contrast, the upstream promoter elements of *pCMPLuc* resulted in only 1.8-fold and 2.6-fold increase of the luciferase activity in CEF and HDM cultures. This indicates that upstream promoter elements can highly increase the tissue- and stage-specific activity of the matrilin-1 promoter.

To line up with the data from genomic footprinting and *in vitro* binding assays, the effect of mutations within the inverted Sox motifs of the conserved Pe1 and Ine support the conclusion that the identified Sox sites were functionally relevant. This indicates important role for Sox family proteins in the constitutive activity of the short matrilin-1 promoter, and also suggests a complex interplay between Sox factors bound to Pe1 and Ine in chondrocytes.

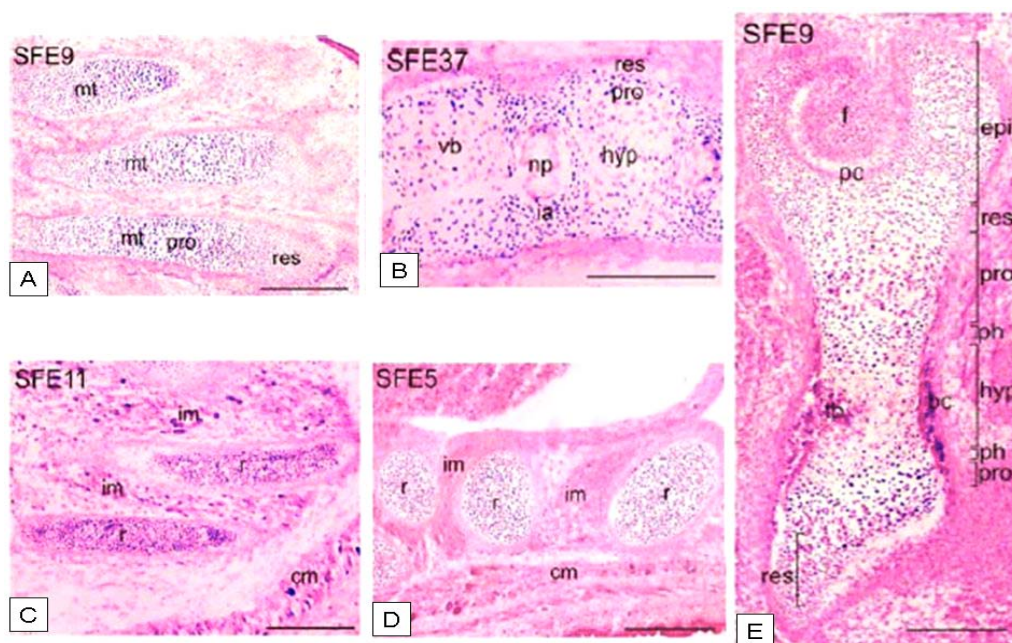
## DISCUSSION

Extracellular matrix metabolism plays a central role in the development of skeletal tissues and in most orthopaedic diseases and trauma such as fracture or osteotomy repair, arthritis, cartilage repair, and congenital skeletal deformity. During development or disease, specific genes must be expressed in order to make or repair appropriate extracellular matrix. For example, specific gene expression patterns are characteristic of bone and cartilage. The precise expression pattern depends on a balance of positive and negative transcription factors, proteins that control the synthesis of mRNA from the specific gene. In cartilage, a number of studies indicate that Sox transcription factors are critical positive regulators in genes such as *Col2a1*, *Col9a2*, *Col11a2*, *Agc1*, and *CD-Rap* [29,133,136,137]. Sox9 has been demonstrated to regulate expression of *Col2a1* via several AGAACAATGG motifs [149]. In addition, negative regulators are also essential to fine tune gene regulation in chondrocytes and to turn off gene expression in noncartilaginous tissues.

The unique feature of the matrilin-1 gene among cartilage-specific genes is the characteristic expression pattern restricted to distinct zones of the growth plate *in vivo* or developmental stages in tissue cultures. Recent analysis of the major regulatory regions of the chicken matrilin-1 gene in transgenic mice revealed that the long promoter alone or in combination with the intronic enhancer as well as the short promoter with the intronic enhancer restricted the transgene expression to the columnar proliferative chondroblasts and prehypertrophic chondrocytes [140]. Transgenic experiments with the chicken matrilin-1 promoter in a mouse have indicated that the tissue-specific control elements are divided between the promoter upstream and intronic regions in a manner similar to that of the *Col11a2* gene [132]. Moreover, NAD1 founder embryos expressing the LacZ reporter gene under the control of the short matrilin-1 promoter that was shared by all above transgenes, also exhibited transgene expression in the developing cartilaginous elements of the chondrocranium, appendicular and axial skeleton (Figure 16), indicating that, the short promoter indeed harbors DNA-elements, which can direct the developmental stage-specific expression of the transgene in chondrocytes [152].

In the work presented in this dissertation, my colleagues and I have taken advantage of the recent advances in regulation of tissue-specific gene expression of matrix proteins in cartilage via Sox

proteins that have been shown to be activating transcription factors indispensable for chondrogenesis. We tried to focus in this study: 1) allocation of *cis*-regulatory elements that are important for the transcriptional regulation of the matrilin-1 gene, in particular those elements which possibly function via interacting with Sox-family proteins in the short promoter region; 2) application of *in vitro* and *in vivo* techniques in order to study the function of these elements in the transcriptional regulation of the chicken matrilin-1 gene. This work has provided valuable insight into the function of these elements *in vivo*, in addition to opening many doors to future experimentation.



**Figure 16. Founder embryos expressing NAD1 (short promoter) transgene** LacZ expression is seen in the developing metatarsals (mt) (A), in vertebral bodies (vb), inner annulus (ia) and nucleus pulposus (np) of developing intervertebral discs (B), in the longitudinal (K) and cross (L) sections of the developing ribs (r) and iliac bone (F). The intensity of X-gal staining is relatively high in the zones of columnar proliferating (pro) chondroblasts and prehypertrophic (ph) chondrocytes, and weaker in the epiphyseal (epi) chondroblasts, resting (res) and hypertrophic (hyp) zones of growth plate cartilage (B,E). SFE, sacrificed founder embryo; im, intervertebral; cm, cutaneous muscles; Scale bar, 200  $\mu$ m.

As a result, we found two *cis*-acting Sox-specific islands (labelled as Pe1 and Ine) within the short promoter of the chicken *matrilin-1*. The two elements directly interact *in vitro* with Sox9, L-Sox5, and Sox6. We also show that, both elements display tissue-specific footprints *in vivo*, and these *cis*-acting elements respond in a cell type-specific manner. Furthermore, the parallel finding by E.



Dpe1 harbors motifs sharing 6/8 bp identity with the preferred recognition sequence for L-Sox5/Sox6 [149]. Dpe2 includes the fully conserved GACACAGAGAA motif, which does not match to any consensus motif of known transcription factors of the TRANSFAC® database (<http://www.gene-regulation.com/pub/databases.html#transfac>). The degree of sequence similarity around the TATA motif was relatively weak as compared to other eukaryotic promoters.

The observation that, contrary to the sequence divergence in other parts of the matrilin-1 regulatory regions in the various species, Pe1 remained strongly conserved under evolutionary pressure between chicken and mammals implies that the element performs a very important function in the transcriptional regulation of the matrilin-1 gene in amniotes.

Phylogenetic conservation of important regulatory sequences has also been reported in other systems, leading to a similar conclusion. For example, *Sox2* enhancers were functionally identified within extragenic sequence blocks clearly conserved between chicken and mammals [153]. From the 25 conserved sequence blocks, however, only two occurred in the fish genome as well, but similarly to our findings reported in the present study, the conserved regulatory elements were hidden within longer stretches of sequence similarity, when only mammalian species were compared. Species-specific variations in the occurrence of conserved regulatory *Sox2* sequence blocks were related to distinct spatio-temporal differences in the gene expression between vertebrate species [154,155]. These and our findings support the conclusion that sequence conservations between chicken and mammals are reliable indications of important regulatory regions within a genetic locus.

Tissue-specific control elements have been identified in the promoter of certain eukaryotic genes, for example between 15 and 200 bp upstream of the TATA box of liver- or osteoblast-specific genes [156-158]. Furthermore, it has also been reported that SOX9 interacting with the partner transcription factor SF-1 recognizes a conserved DNA element approximately 100 bp upstream of the TATA box of mammalian anti-Müllerian hormone genes [124]. Even though inverted Sox motifs are known to play an essential role in the function of chondrocyte-specific enhancers of the *Col2a1* and *Col11a2* genes, conserved blocks similar to Pe1 could be found neither in their proximal promoter elements nor within the 3 kb region of the putative promoters of these genes in human, mouse and rat by computer search using the programs DIALIGN2 and MEME [144,145]. In accordance with this

observation, the proximal promoter region of neither the *Col2a1* and *Coll1a1* nor the *Agc* and the *CD-Rap* genes were reported to mediate cartilage-specific regulation [132,159,160]. Interestingly, consistent with its conserved position in amniotes, the Pe1 element seems to function at a certain distance from the TATA box. In this respect, it also clearly differs from the cartilage-specific enhancer elements of other cartilage-protein genes [29,132,136,137,150,159,160]. These findings imply that Pe1 is a unique control element of the matrilin-1 gene not shared by other cartilage protein genes.

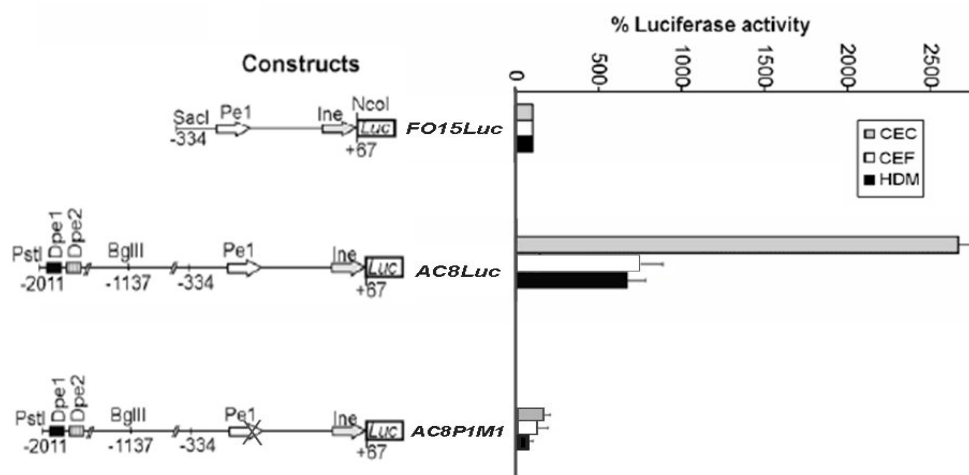
The sequence homology among amniotes within Ine was restricted only to a short motif, but two pairs of inverted Sox-binding sites were identified in chicken (Figure 6). Besides being recognized by antibodies of Sox9, L-Sox5 and Sox6 *in vitro*, genomic footprinting also revealed the occupancy of these motifs by transcription factors bound in chondrocytes *in vivo*. Transient expression studies confirmed that both Pe1 and Ine elements significantly contribute to the moderate activity of the short promoter in chondrocytes. Based on these data, we hypothesize that the element does not drive the high cartilage-specific expression of the promoter as an enhancer, but may rather act by modulating the promoter activity and mediating the effect of distal promoter and intronic enhancer elements. We also hypothesize that distal promoter and intronic elements may also function by forming multi-protein complexes via interacting with Pe1 and Ine.

Hypothesis proposed here are consistent with our previous observations [138-140]. TR70 transgenic mice expressing LacZ reporter gene under 2011 bp promoter control region of the chicken matrilin-1 along with histological analysis confirmed developmental stage-specific X-gal staining in cartilage, congruent with that of the endogenous *Matn1* gene in all developing skeletal elements. The activity of the transgenes was restricted to the columnar proliferating and prehypertrophic chondrocytes, and no X-gal staining was detected in the condensed mesenchyme or precartilaginous cells. On the contrary, histological analysis of the founder embryos expressing 338 bp short promoters alone in NAD1 transgenes, also exhibited uniform X-gal staining in the cartilaginous elements of the developing chondrocranium, appendicular and axial skeleton, implying that the short promoter indeed harbors chondrocyte-specific DNA-elements [152]. The transgene activity, however, was relatively low as compared to TR70 and no zonal differences were seen in the expression pattern suggesting that the DNA-elements responsible for the developmental stage-specific activity of the gene must be



located outside of the short promoter. Consistently, when longer promoter fragment harboring the putative Dpe1 and Dpe2 elements were tested in transient expression assays [152], we observed 26.9-fold enhancement of the reporter gene activity of construct *AC8Luc* (made by A. Daraba) as opposing to *FO15Luc* in CEC cultures (Figure 18). Comparison of the reporter gene activity between *FO15Luc* and *AC8Luc*, in transient expression and between the NAD1, TR70 and VAM1 transgenes [140,152] stress that distal promoter and intronic elements highly enhance the promoter activity and greatly increase the zonal differences in the transgene expression.

Although Pe1 can bind to Sox9, L-Sox5 and Sox6 *in vitro*, it may show a preference for Sox9 *in vivo*, as the inverted Sox motifs carried by Pe1 are more similar to the preferred binding sites of Sox9, than to those of L-Sox5/Sox6. On the other hand, the conserved Dpe1 element carries motifs more similar to the preferred binding sites of L-Sox5/Sox6. Binding of Dpe1 to these factors and the interaction with nucleoprotein complexes formed on Pe1 may be necessary for the high tissue- and developmental stage-specific activity of the 2011 bp promoter in *AC8Luc* and TR70 transgenic mice [140,152]. This hypothesis is supported by recent observation from our laboratory demonstrating that point mutations in Pe1 highly reduce the activity of the long promoter in AC8 (Figure 18) (personal communication with E. Kenesi, S. T. Oommen, A. Nagy and I. Kiss).



**Figure 18. Functional analysis of long matrilin-1 promoter (*AC8Luc*) and the role of Pe1 in transient expression assay** Point mutations of Pe1M1, were introduced into *AC8Luc* in derivative *AC8P1M1* (construct was made by A. Nagy). Luciferase activities are expressed in % of that of the *FO15Luc* set at 100%. Values represent the averages  $\pm$  S.E.M. of 2-5 independent transfection experiments in CEC, CEF and HDM cultures.

These data indicate that Sox9 and L-Sox5/Sox6-binding sites may be separated over a large distance in the regulatory region of the matrilin-1 gene. The importance of L-Sox5/Sox6 binding is supported by the observation that Sox9 alone is not sufficient for the activation of the matrilin-1 gene in the absence of Sox5 and Sox6 proteins [28].

In our experiments, GST–Sox9 formed only a single complex with Pe1 and Ine. Even though Sox9 is not capable of forming homodimers in solution [29,161,162], consistent with observations from other laboratories [150,163], we found that it could bind only to intact inverted pairs of Sox motifs (Figure 11B), thus supporting the conclusion that Sox9 dimerization might have occurred upon DNA binding. As opposed to this, Sox9 was reported to bind as a monomer to *cis* elements involved in sex determination [163].

This is the first report demonstrating that paired Sox-specific sites mapped close to the TATA motif can play a functional role in the transcriptional activation of a cartilage protein gene. The close proximity of the Sox-binding sites of Ine to the TATA box raises questions regarding the possible interactions of Sox proteins with general transcription machinery. The involvement of HMG1 domain proteins in the regulation of gene transcription has been suspected for a long time, and conflicting effects have been published. Depending on the different systems and conditions used, either repression or activation of transcription has been observed *in vitro* [164,165]. It has been reported for example that the abundant HMG1 protein, which binds angled structures in the DNA without any sequence specificity, can interact in a species-specific manner with the core domain of the TATA box-binding protein (TBP), and block the formation of the preinitiation complex (PIC) by preventing the binding of TFIIB to TBP [164,166]. Such repression of RNA polymerase II transcription can be reversed by TFIIA. Differing from the HMG1 protein, Sox9 binds to the DNA in a sequence-specific manner and has a transactivation domain. We hypothesize that, according to the intrinsic flexibility of DNA bending and cooperative binding ability with multiple transcription factors, Sox proteins (L-Sox5, Sox6 and Sox9) may interact with the components of the general transcription factors on Ine, which may generate a surface complementary to the RNA polymerase II transcription machinery. It is yet to be confirmed that this interaction or combination of general transcription machinery really function on the natural Ine element of the matrilin-1 gene, and the interacting partners are still to be found. In any

case, our results suggest that the mapped Sox sites in Ine might be biologically relevant for the possible modulation of PIC formation *in vivo*.

By bending the DNA, HMG-box proteins are known to promote the binding of other transcription factors to the DNA. Lining up with these observations, our results suggest that, in addition to Sox proteins, other transcription factors may also be involved in the activity of the short matrilin-1 promoter. Sox9 is known to interact with a number of partner factors, including steroidogenic factor-1 (SF-1) and long cell lineage-specific musculoaponeurotic fibrosarcoma (Lc-Maf) [29,122,124]. One explanation for the conserved position of the Pe1 element can be that it may function by bending the DNA and promoting the interaction between the components of the polymerase II transcription machinery and ubiquitous factors bound to the proximal promoter elements. This assumption is supported by the observation that the short promoter includes several putative binding sites for ubiquitous transcription factors, such as Sp1 and NFI, and based on our genomic footprinting studies, these motifs were also occupied by transcription factors bound *in vivo* in chondrocytes (Figure 7; and results not shown). Furthermore, Sox9 may interact with different partner factors during subsequent steps of chondrogenesis, thereby contributing to the developmental stage-specific activity of the matrilin-1 gene. However, future studies will be needed for mapping the putative cofactor-binding sites and identifying the interacting Sox partner factors on the Pe1 and Ine elements.

Taken together, our data suggest that a) the cartilage-specific control elements are dispersed in the matrilin-1 regulatory regions and b) modularly arranged cartilage- and neural crest-specific enhancer and silencer elements located in the promoter upstream and intronic control regions regulate the expression of the chicken matrilin-1 gene during ontogenetic development. This hypothesis lines up with our previous observation, where *in vitro* experiments confirmed the presence of the two promoter proximal silencer elements, SI and SII [139]. SII resides right next to the Pe1, and SI is located within 20 bp upstream of the TATA motif. Furthermore, we have identified recently cis-elements containing Sox9-binding sites in the first intron as well (Rauch et al., to be published elsewhere).

To allow tissue-specific expression, genes are negatively regulated by activation of negative regulators coupled with repression of positive factors where the genes should not be expressed. A number of negative regulators for cartilage-specific matrix proteins have been identified recently. The promoter region of *Col2a1* contains several E-boxes (CANNTG, which binds basic helix-loop-helix transcription factors), and  $\Delta$ EF1 is proposed to be a repressor functioning on the E-boxes.  $\Delta$ EF1 is reciprocally expressed compared with *Col2a1*; overexpression of estrogen-inducible transcription factor delta ( $\Delta$ EF1) represses expression of reporter construct containing a *Col2a1* promoter/enhancer [167].  $\Delta$ EF1 null mutant mice exhibit various skeletal abnormalities [168]. Mouse Snail and Slug are family members of zinc finger transcription factors, are also reported to bind to the proximal E-box of *Col2a1* promoter [169]. Overexpression of Snail represses the expression of the reporter gene as well as endogenous expression of *Col2a1* and aggrecan in chondrogenic progenitor cells ATDC5. Therefore, the promoter region of *Col2a1* contains several potent functional elements for tissue specific expression. The other important DNA motif in *Col2a1* is located in the first intron [123,170]. A study of transgenic mice revealed that the minimum enhancer element was able to direct cartilage-specific expression of the reporter gene in vivo. The element contains functional HMG-like sites, which are targets of Sox proteins, thus turning on the gene in Sox9-expressing cells. However, the element also contains a silencer domain to turn off in many other tissues. In fact,  $\alpha$ A-Crystallin binding protein-1 (CYRBP1), a zinc finger transcription factor, binds next to the HMG-like site, competing with Sox9 binding within the control element of the first intron of *Col2a1*, resulting potentially in repression of gene transcription [171]. A similar model whereby negative regulators and positive HMG-like sites in a tissue-specific control element is also recognized in the upstream promoter of *Col11a2*. The study of transgenic mice revealed that the -530 bp promoter directed cartilage-specific expression in vivo, whereas deletion to -453 bp resulted in loss of specificity [172]. This tissue-specific control fragment from -530 bp to -453 bp contains multiple HMG-like sites and one of them is functional for Sox9 [133]. A negative regulator, NT2, a Krüppel-associated box-zinc finger protein, was identified as binding within the element and represses the promoter activity [173].

Accordingly, genes are regulated by a balance of positive and negative factors. It is not surprising therefore those tissue-specific genes are regulated by a few critical positive factors and many negative regulators in order to exhibit a strictly restricted expression.

## SUMMARY OF NOVEL FINDINGS

1. Throughout the 1.2-kb promoter region tested between positions -1137 and +64, we identified two potential Sox-specific islands labelled as Pe1 and Ine in the short promoter of chicken matrilin-1 gene.
  - Pe1 element includes a pair of inverted motifs highly similar to the AGAACAATGG motif, which was shown to be the preferred binding site of Sox9 *in vitro*.
  - Ine1 harbors two pairs of inverted putative Sox motifs with 5/10, 6/10, 7/10 and 4/9 identity with the consensus Sox-binding sites CA/TTTGA/TA/T *in vitro*.
2. We observed *in vivo* occupancy of the Sox motifs in genomic footprinting in the expressing cell type, but not in the non-expressing, which support the involvement of Pe1 and Ine in the tissue-specific regulation of the gene. Apart from Pe1 and Ine, in chondrocytes, footprints were also visible at the NFI contact points of the SI and SII elements that were identified previously by Szabó et al in 1995.
3. We provided evidence that both elements (Pe1 and Ine) interact with chondrogenic transcription factors of Sox9, L-Sox5 and Sox6 *in vitro*.
  - Eventhough, Pe1 contain pair of inverted Sox9 motifs, we noticed that the element form only a single nucleoprotein complex with GST–Sox9. Point mutations in either Sox motifs interfered with this complex, suggesting that GST-Sox9 could bind only to intact inverted pairs of Sox motifs and that Sox9 dimerization might have occurred upon DNA binding.
4. Using transient expression studies, we confirmed that both Pe1 and Ine elements significantly contribute to the moderate activity of the short promoter in chondrocytes. We hypothesize that the element does not drive the high cartilage-specific expression of the promoter as an enhancer, but

may rather act by modulating the promoter activity and mediating the effect of distal promoter and intronic enhancer elements.

5. This is the first report demonstrating that paired Sox-specific sites mapped within Ine, that is close to the TATA motif can play a functional role in the transcriptional activation of a cartilage protein gene. We hypothesize that, according to the intrinsic flexibility of DNA bending and cooperative binding ability with multiple transcription factors, Sox proteins (L-Sox5, Sox6 and Sox9) may interact with the components of the general transcription machinery assembled on Ine, which may generate a surface complementary to the RNA polymerase II transcription machinery. In addition, distal promoter and intronic elements may also function by forming multi-protein complexes via interacting with Pe1 and Ine.

These findings imply for the first time an important role of cis-regulatory elements that are functioning via Sox family proteins in the transcriptional upregulation of the chicken matrilin-1 gene.

## ACKNOWLEDGEMENTS

This work was carried out in the Institute of Biochemistry, Biological Research Center of Hungarian Academy of Sciences, Szeged, Hungary, during the years 1999-2005. I wish to express my deepest gratitude to my supervisor, Dr. Ibolya Kiss, Ph.D. for giving me the opportunity to work in her excellent group and for introducing me to the field of molecular biology and matrix biology. The depth of her contribution to these studies has been widespread and is too expansive to list here. Her guidance and enthusiastic support during each step of the process have been of great benefit to the present work. I wish to express my appreciation and respect for Dr. Ferenc Deák for his never-ending enthusiasm and for putting to good use a sharp, inquiring mind.

I would also like to acknowledge the Szeged Biological Research Center, especially past and present Directors of the Institute of Biochemistry, for providing financial support as well as for arranging and conducting useful courses and seminars.

I also wish to express my thanks to all the members of the Connective Tissue Laboratory, past and present. In particular, I wish to thank to Éva Korpos, Ildikó Karcagi, and Oana Sicora, for their invaluable support, love and friendship during these years. I would like to extend special thanks for the excellent technical support that I have received from Aniko Simon, Ildikó Kravjár, Katalin Kovács and Irén Fekete. Their unequalled skills at the bench have been an integral part of much of the work presented in this thesis. Furthermore, Ildikó Karcagi, Tibor Rauch, Andrea Nagy, Ildikó Sinko, Andreea Daraba, and Endre Barta are warmly acknowledged for their contributions to exploring the importance of the matrilin-1, which formed the essential basis of the work presented here. To Tibor Rauch and András Blastyák I owe the most gratitude, who has enthusiastically taught their experience and expertise in the protein-DNA interaction techniques. Enormous thanks to Mária Tóth for providing of quality art work. I am also especially indebted to Piroska Szabó who introduced me into Genomic footprinting and Gene manipulation via Homologous recombination techniques.

I would not have enjoyed life in Szeged so much without my dear friends from ITC. Their companionship has provided the most important and vital meaning in the midst of this furious scientific whirlwind.

Finally, I would like to express my thoughts and deepest gratitude to all of my family for their unconditional support and encouragement. My mother Myadagmaa contributed greatly on this. I thank her for everything from A to Z. My children, Belgutei and Enkhluun, have been my greatest source of inspiration. Being tired of missing them, missing their fantastic stories, smiles, their little voices, being tired of not being able to fulfill their requests are the best motivations I know for finishing this work. This thesis would not have been possible without them.



## REFERENCES

1. Olsen, B. R., Reginato, A. M. and Wang, W. (2000) Bone development. *Annu Rev Cell Dev Biol.* **16**, 191-220. Review
2. So, C. L., Kaluarachchi, K., Tam, P. P. and Cheah, K. S. (2001) Impact of mutations of cartilage matrix genes on matrix structure, gene activity and chondrogenesis. *Osteoarthritis Cartilage* **9** Suppl **A**, S160–S173
3. Kirn-Safran, C. B., Gomes, R. R., Brown, A. J. and Carson, D. D. (2004) Heparan sulfate proteoglycans: coordinators of multiple signaling pathways during chondrogenesis. *Birth Defects Res C Embryo Today* **72**, 69–88
4. Mundlos, S. and Olsen, B. R. (1997) Heritable diseases of the skeleton. Part I: Molecular insights into skeletal development-transcription factors and signaling pathways. *FASEB J.* **11**, 125-132
5. Mundlos, S. and Olsen, B. R. (1997) Heritable diseases of the skeleton. Part II: Molecular insights into skeletal development-matrix components and their homeostasis. *FASEB J.* **11**, 227-233
6. Erlebacher, A., Filvaroff, E. H., Gitelman, S. E. and Derynck, R. (1995) Toward a molecular understanding of skeletal development. *Cell* **80**, 271–278
7. Kronenberg, H. M. (2003) Developmental regulation of the growth plate. *Nature* **423**, 332–336
8. Zelzer, E. and Olsen, B. R. (2003) The genetic basis for skeletal diseases. *Nature* **423**, 343–348
9. Karsenty, G. (2003) The complexities of skeletal biology. *Nature* **423**, 316–318
10. Hunziker, E. B. (1994) Mechanism of longitudinal bone growth and its regulation by growth plate chondrocytes. *Microscopy research and technique* **28**, 505-519
11. Hansson, L. I. (1967) Daily growth in length of diaphysis measured by oxytetracycline in rabbit normally and after medullary plugging. *Acta Orthop Scand.* Suppl 101:1
12. Stevenson, S., Hunziker, E. B., Herrmann, W. and Schenk, R. K. (1990) Is longitudinal bone growth influenced by diurnal variation in the mitotic activity of chondrocytes of the growth plate? *J Orthop Res.* **8**, 132-135
13. Aszodi, A., Modis, L., Paldi, A., Rencendorj, A., Kiss, I. and Bosze, Z. (1994) The zonal expression of chicken cartilage matrix protein gene in the developing skeleton of transgenic mice. *Matrix Biol.* **14**, 181-90
14. Iannotti, J. P. (1990) Growth plate physiology and pathology. *Orthop Clin North Am.* **21**, 1-17. Review
15. Breur, G. J., VanEnkevort, B. A., Farnum, C. E. and Wilsman, N. J. (1991) Linear relationship between the volume of hypertrophic chondrocytes and the rate of longitudinal bone growth in growth plates. *J Orthop Res.* **9**, 348-59

16. Muratoglu, S., Bachrati, C., Malpeli, M., Szabó, P., Neri, M., Dozin, B., Deák, F., Cancedda, R. and Kiss, I. (1995) Expression of the cartilage matrix protein gene at different chondrocyte developmental stages. *Eur J Cell Biol.* **68**, 411-8
17. de Crombrughe, B., Lefebvre, V. and Nakashima, K. (2001) Regulatory mechanisms in the pathways of cartilage and bone formation. *Curr. Opin. Cell Biol.* **13**, 721–727.
18. Lefebvre, V. (2002) Toward understanding the functions of the two highly related Sox5 and Sox6 genes. *J. Bone Miner. Metab.* **20**, 121–130
19. de Crombrughe, B., Lefebvre, V., Behringer, R. R., Bi, W., Murakami, S. and Huang, W. (2000) Transcriptional mechanisms of chondrocyte differentiation. *Matrix Biol.* **19**, 389-94. Review
20. Lefebvre, V. and de Crombrughe, B. (1998) Toward understanding SOX9 function in chondrocyte differentiation. *Matrix Biol.* **16**, 529-40
21. Bi, W., Deng, J. M., Zhang, Z., Behringer, R. R. and de Crombrughe, B. (1999) Sox9 is required for cartilage formation. *Nat Genet.* **22**, 85-9
22. Wagner, T., Wirth, J., Meyer, J., et al. (1994) Autosomal sex reversal and campomelic dysplasia are caused by mutations in and around the SRY-related gene SOX9. *Cell* **79**, 1111–1120
23. Foster, J. W., Dominguez-Steglich, M. A., Guioli, S., et al. (1994) Campomelic dysplasia and autosomal sex reversal caused by mutations in an SRY-related gene. *Nature* **372**, 525–530
24. Ng, L. J., Wheatley, S., Muscat, G. E., et al. (1997) SOX9 binds DNA, activates transcription, and coexpresses with type II collagen during chondrogenesis in the mouse. *Dev Biol* **183**, 108–121
25. Akiyama, H., Chaboissier, M. C., Martin, J. F., Schedl, A. and de Crombrughe, B. (2002) The transcription factor Sox9 has essential roles in successive steps of the chondrocyte differentiation pathway and is required for expression of Sox5 and Sox6. *Genes Dev* **16**, 2813–2828
26. Akiyama, H., Lyons, J. P., Mori-Akiyama, Y., et al. (2004) Interactions between Sox9 and beta-catenin control chondrocyte differentiation. *Genes Dev* **18**, 1072– 1087
27. Bi, W., Huang, W., Whitworth, D. J., Deng, J. M., Zhang, Z., Behringer, R. R. and de Crombrughe, B. (2001) Haploinsufficiency of Sox9 results in defective cartilage primordia and premature skeletal mineralization. *Proc Natl Acad Sci U S A.* **98**, 6698-6703
28. Smits, P., Li, P., Mandel, J., Zhang, Z., Deng, J. M., Behringer, R. R., de Crombrughe, B. and Lefebvre, V. (2001) The transcription factors L-Sox5 and Sox6 are essential for cartilage formation. *Dev Cell.* **1**, 277-90
29. Lefebvre, V., Li, P. and de Crombrughe, B. (1998) A new long form of Sox5 (L-Sox5), Sox6 and Sox9 are coexpressed in chondrogenesis and cooperatively activate the type II collagen gene. *EMBO J.* **17**, 5718-33
30. Bell, D. M., Leung, K. K., Wheatley, S. C., Ng, L. J., Zhou, S., Ling, K. W., Sham, M. H., Koopman, P., Tam, P. P. and Cheah, K. S. (1997) SOX9 directly regulates the type-II collagen gene. *Nat Genet.* **16**, 174-178

31. Lefebvre, V., Huang, W., Harley, V. R., Goodfellow, P. N. and de Crombrughe, B. (1997) SOX9 is a potent activator of the chondrocyte-specific enhancer of the pro alpha1(II) collagen gene. *Mol Cell Biol* **17**, 2336-2346
32. Kronenberg, H. M. (2003) Developmental regulation of the growth plate. *Nature* **423**, 332–336
33. Pizette, S. and Niswander, L. (2000) BMPs are required at two steps of limb chondrogenesis: formation of prechondrogenic condensations and their differentiation into chondrocytes. *Dev Biol.* **219**, 237-49
34. Brunet, L. J., McMahon, J. A., McMahon, A. P. and Harland, R. M. (1998) Noggin, cartilage morphogenesis, and joint formation in the mammalian skeleton. *Science.* **280**, 1455-7
35. Tickle C. (2003) Patterning systems from one end of the limb to the other. *Dev Cell.* **4**, 449-58. Review
36. Minina, E., Kreschel, C., Naski, M. C., Ornitz, D. M. and Vortkamp, A. (2002) Interaction of FGF, Ihh/Pthlh, and BMP signaling integrates chondrocyte proliferation and hypertrophic differentiation. *Dev Cell.* **3**, 439-49
37. Minina, E., Wenzel, H. M., Kreschel, C., Karp, S., Gaffield, W., McMahon, A. P. and Vortkamp, A. (2001) BMP and Ihh/PTHrP signaling interact to coordinate chondrocyte proliferation and differentiation. *Development*, **128**, 4523-34
38. Vortkamp, A., Lee, K., Lanske, B., Segre, G. V., Kronenberg, H. M. and Tabin, C. J. (1996) Regulation of rate of cartilage differentiation by Indian hedgehog and PTH-related protein. *Science*, **273**, 613-622
39. St-Jacques, B., Hammerschmidt, M. and McMahon, A. P. (1999) Indian hedgehog signaling regulates proliferation and differentiation of chondrocytes and is essential for bone formation. *Genes Dev*, **13**, 2072-2086
40. Ornitz, D. M. and Marie, P. J. (2002) FGF signaling pathways in endochondral and intramembranous bone development and human genetic disease. *Genes Dev.* **16**, 1446–1465
41. Sahni, M., Ambrosetti, D. C., Mansukhani, A., Gertner, R., Levy, D. and Basilico, C. (1999) FGF signaling inhibits chondrocyte proliferation and regulates bone development through the STAT-1 pathway. *Genes Dev*, **13**, 1361-1366
42. Su, W. C., Kitagawa, M., Xue, N., Xie, B., Garofalo, S., Cho, J., Deng, C., Horton, W. A. and Fu, X. Y. (1997) Activation of Stat1 by mutant fibroblast growth-factor receptor in thanatophoric dysplasia type II dwarfism. *Nature*, **386**, 288-292
43. Bellus, G. A., McIntosh, I., Smith, E. A., Aylsworth, A. S., Kaitila, I., Horton, W. A., Greenhaw, G. A., Hecht, J. T. and Francomano, C. A. (1995) A recurrent mutation in the tyrosine kinase domain of fibroblast growth factor receptor 3 causes hypochondroplasia. *Nat Genet.* **10**, 357-359
44. Rousseau, F., Bonaventure, J., Legeai-Mallet, L., Pelet, A., Rozet, J. M., Maroteaux, P., Le Merrer, M. and Munnich, A. (1994) Mutations in the gene encoding fibroblast growth factor receptor-3 in achondroplasia. *Nature*, **371**, 252-254

45. Rousseau, F., Saugier, P., Le Merrer, M., Munnich, A., Delezoide, A. L., Maroteaux, P., Bonaventure, J., Narcy, F. and Sanak, M. (1995) Stop codon FGFR3 mutations in thanatophoric dwarfism type 1. *Nat Genet*, **10**, 11-12
46. Shiang, R., Thompson, L. M., Zhu, Y. Z., Church, D. M., Fielder, T. J., Bocian, M., Winokur, S. T. and Wasmuth, J. J. (1994) Mutations in the transmembrane domain of FGFR3 cause the most common genetic form of dwarfism, achondroplasia. *Cell*, **78**, 335-342
47. Tavormina, P. L., Shiang, R., Thompson, L. M., Zhu, Y. Z., Wilkin, D. J., Lachman, R. S., Wilcox, W. R., Rimoin, D. L., Cohn, D. H. and Wasmuth, J. J. (1995) Thanatophoric dysplasia (types I and II) caused by distinct mutations in fibroblast growth factor receptor 3. *Nat Genet*, **9**, 321-328
48. Francomano, C. A. and Deng, C. X. (2000) A neonatal lethal mutation in FGFR3 uncouples proliferation and differentiation of growth plate chondrocytes in embryos. *Hum Mol Genet*, **9**, 1603-1613
49. Iwata, T., Li, C. L., Deng, C. X. and Francomano, C. A. (2001) Highly activated Fgfr3 with the K644M mutation causes prolonged survival in severe dwarf mice. *Hum Mol Genet*, **10**, 1255-64
50. Li, C., Chen, L., Iwata, T., Kitagawa, M., Fu, X. Y. and Deng, C. X. (1999) A Lys644Glu substitution in fibroblast growth factor receptor 3 (FGFR3) causes dwarfism in mice by activation of STATs and ink4 cell cycle inhibitors. *Hum Mol Genet*, **8**, 35-44
51. Naski, M. C., Colvin, J. S., Coffin, J. D. and Ornitz, D. M. (1998) Repression of hedgehog signaling and BMP4 expression in growth plate cartilage by fibroblast growth factor receptor 3. *Development*, **125**, 4977-4988
52. Chen, L., Li, C., Qiao, W., Xu, X. and Deng, C. (2001) A Ser(365)→Cys mutation of fibroblast growth factor receptor 3 in mouse downregulates Ihh/PTHrP signals and causes severe achondroplasia. *Hum Mol Genet*, **10**, 457-465
53. Argraves, W. S., Deak, F., Sparks, K. J., Kiss, I. and Goetinck, P. F. (1987) Structural features of cartilage matrix protein deduced from cDNA. *Proc Natl Acad Sci U S A*. **84**, 464-8
54. Kiss, I., Deák, F., Holloway, R. G., Delius, H., Mebust, K. A., Frimberger, E., Argraves, W. S., Tsonis, P. A., Winterbottom, N. and Goetinck, P. F. (1989) Structure of the gene for cartilage matrix protein, a modular protein of the extracellular matrix. Exon/intron organization, unusual splice sites, and relation to alpha chains of beta 2 integrins, von Willebrand factor, complement factors B and C2, and epidermal growth factor. *J. Biol. Chem.* **264**, 8126-8134
55. Hauser, N. and Paulsson, M. (1994) Native cartilage matrix protein (CMP): a compact trimer of subunits assembled via a coiled-coil alpha-helix. *J Biol Chem.* **269**, 25747-53
56. Deák, F., Wagener, R., Kiss, I. and Paulsson, M. (1999) The matrilins: a novel family of oligomeric extracellular matrix proteins. *Matrix Biol.* **18**, 55-64

57. Deák, F., Piecha, D., Bachrati, C., Paulsson, M., and Kiss, I. (1997) Primary structure and expression of matrilin-2, the closest relative of cartilage matrix protein within the von Willebrand factor type A module superfamily. *J. Biol. Chem.* **272**, 9268-9274
58. Muratoglu, S., Krysan, K., Balazs, M., Sheng, H., Zakany, R., Modis, L., Kiss, I. and Deak, F. (2000) Primary structure of human matrilin-2, chromosome location of the MATN2 gene and conservation of an AT-AC intron in matrilin genes. *Cytogenet Cell Genet.* **90**, 323-7
59. Belluoccio, D. and Trueb, B. (1997) Matrilin-3 from chicken cartilage. *FEBS Lett.* **415**, 212-216
60. Piecha, D., Muratoglu, S., Mörgelin, M., Hauser, N., Studer, D., Kiss, I., Paulsson, M. and Deák, F. (1994) Matrilin-2, a large, oligomeric matrix protein, is expressed by a great variety of cells and forms fibrillar networks. *J. Biol. Chem.* **274**, 13353-13361
61. Wagener, R., Kobbe, B. and Paulsson, M. (1997) Primary structure of matrilin-3, a new member of a family of extracellular matrix proteins related to cartilage matrix protein (matrilin 1) and von Willebrand factor. *FEBS Lett.* **413**, 129-134
62. Wagener, R., Kobbe, B. and Paulsson, M. (1998) Matrilin-4, a new member of the family of extracellular matrix proteins. *FEBS Lett.* **436**, 123-127
63. Segat, D., Nitsche, P. D., Klatt, A. R., Piecha, D., Korpos, E., Deák, F., Wagener, R., Paulsson, M. and Smyth, N. (2000) Expression of matrilin-1, -2 and -3 in developing mouse limbs and heart. *Matrix Biol.* **19**, 649-655
64. Chapman, K. L., Mortier, G. R., Chapman, K., Loughlin, J., Grant, M. E., and Briggs, M. D. (2001) Mutations in the region encoding the von Willebrand factor A domain of matrilin-3 are associated with multiple epiphyseal dysplasia. *Nat. Genet.* **28**, 393-396
65. Whittaker, C. A. and Hynes, R. O. (2002) Distribution and evolution of von Willebrand/Integrin A domains: widely dispersed domains with roles in cell adhesion and elsewhere. *Mol. Biol. Cell* **13**: 3369-3387.
66. Wu, J. J. and Eyre, D. R. (1998) Matrilin-3 forms disul.de-linked 490 oligomers with matrilin-1 in bovine epiphyseal cartilage. *J. Biol. Chem.* **273**, 17433-17438
67. Frank, S., Schulthess, T., Landwehr, R., Lustig, A., Mini, T., Jenö, P., Engel, J. and Kammerer, R.A. (2002) Characterization of the matrilin coiled-coil domains reveals seven novel isoforms. *J. Biol. Chem.* **277**, 19071–19079
68. Winterbottom, N., Tondravi, M. M., Harrington, T. L., Klier, F. G., Vertel, B. M., and Goetinck, P. F. (1992) Cartilage matrix protein is a component of the collagen fibril of cartilage. *Dev. Dyn.* **193**, 266-276
69. Chen, Q., Johnson, D. M., Haudenschild, D. R., Tondravi, M. M., and Goetinck, P. F. (1995) Cartilage matrix protein forms a type II collagen-independent filamentous network: analysis in primary cell cultures with a retrovirus expression system. *Mol. Biol. Cell* **6**, 1743-1753

70. Chen, Q., Zhang, Y., Johnson, D. M., and Goetinck, P. F. (1999) Assembly of a novel cartilage matrix protein filamentous network: molecular basis of differential requirement of von Willebrand factor A domains. *Mol. Biol. Cell* **10**, 2149-2162
71. Klatt, A. R., Nitsche, D. P., Kobbe, B., Morgelin, M., Paulsson, M., and Wagener, R. (2000) Molecular structure and tissue distribution of matrilin-3, a filament-forming extracellular matrix protein expressed during skeletal development. *J. Biol. Chem.* **275**, 3999-4006
72. Klatt, A. R., Nitsche, D. P., Kobbe, B., Macht, M., Paulsson, M., and Wagener, R. (2001) Molecular structure, processing, and tissue distribution of matrilin-4. *J. Biol. Chem.* **276**, 17267-75
73. Hauser, N., Paulsson, M., Heinegard, D. and Mörgelin, M. (1996) Interaction of cartilage matrix protein with aggrecan: Increased covalent cross-linking with tissue maturation. *J. Biol. Chem.* **271**, 32247-32252
74. Makihira, S., Yan, W., Ohno, S., Kawamoto, T., Fujimoto, K., Okimura, A., Yoshida, E., Noshiro, M., Hamada, T. and Kato, Y. (1999) Enhancement of cell adhesion and spreading by a cartilage-specific noncollagenous protein, cartilage matrix protein (CMP/Matrilin-1), via integrin  $\alpha 1\beta 1$ . *J Biol Chem.* **274**, 11417-23
75. Huang, X., Birk, D. E. and Goetinck, P. F. (1999) Mice lacking matrilin-1 (cartilage matrix protein) have alterations in type II collagen fibrillogenesis and fibril organization. *Dev. Dyn.* **216**, 434-441
76. Aszódi, A., Bateman, J. F., Hirsch, E., Baranyi, M., Hunziker, E. B., Hauser, N., Bosze, Z. and Fässler, R. (1999) Normal skeletal development of mice lacking matrilin-1: redundant function of matrilins in cartilage? *Mol. Cell. Biol.* **19**, 7841–7845
77. Buckner, J. H., Wu, J. J., Reife, R. A., Terato, K. and Eyre, D. R. (2000) Autoreactivity against matrilin-1 in a patient with relapsing polychondritis. *Arthritis Rheum.* **43**, 939–943
78. Okimura, A., Okada, Y., Makihira, S., Pan, H., Yu, L., Tanne, K., Imai, K., Yamada, H., Kawamoto, T., Noshiro, M., Yan, W. and Kato, Y. (1997) Enhancement of cartilage matrix protein synthesis in arthritic cartilage. *Arthritis Rheum.* **40**, 1029–1036
79. Saxne, T. and Heinegard, D. (1989) Involvement of nonarticular cartilage, as demonstrated by release of a cartilage-specific protein, in rheumatoid arthritis. *Arthritis Rheum.* **32**, 1080–1086
80. Tsonis, P. A. and Goetinck, P. F. (1988) Expression of cartilage-matrix genes and localization of their translation products in the embryonic chick eye. *Exp Eye Res.* **46**, 753-64
81. Paulsson, M., Inerot, S. and Heinegard, D. (1984) Variation in quantity and extractability of the 148-kilodalton cartilage protein with age. *Biochem J.* **221**, 623-30
82. Muratoglu, S., Bachrati, C., Malpeli, M., Szabó, P., Neri, M., Dozin, B., Deák, F., Cancedda, R. and Kiss, I. (1995) Expression of the cartilage matrix protein gene at different chondrocyte developmental stages. *Eur J Cell Biol.* **68**, 411-8

83. Szuts, V., Möllers, U., Bittner, K., Schürmann, G., Muratoglu, S., Deák, F., Kiss, I. and Bruckner, P. (1998) Terminal differentiation of chondrocytes is arrested at distinct stages identified by their expression repertoire of marker genes. *Matrix Biol.* **17**, 435-448
84. Franzen, A., Heinegard, D. and Solursh, M. (1987) Evidence for sequential appearance of cartilage matrix proteins in developing mouse limbs and in cultures of mouse mesenchymal cells. *Differentiation*, **36**, 199-210
85. Stirpe, N. S. and Goetinck, P. F. (1989) Gene regulation during cartilage differentiation: temporal and spatial expression of link protein and cartilage matrix protein in the developing limb. *Development*, **107**, 23-33
86. Mundlos, S. and Zabel, B. (1994) Developmental expression of human cartilage matrix protein. *Dev. Dyn.* **199**, 241-252
87. Aszodi, A., Hauser, N., Studer, D., Paulsson, M., Hiripi, L., and Bosze, Z. (1996) Cloning, sequencing and expression analysis of mouse cartilage matrix protein cDNA. *Eur. J. Biochem.* **236**, 970-977
88. Cancedda, R., Cancedda, F. D. and Castagnola, P. (1995) Chondrocyte differentiation. *Int Rev Cytol.* **159**, 265-358
89. Smale, S. T. (1994) Core promoter architecture for eukaryotic protein-coding genes. In *Transcription. Mechanisms and Regulation*, pp. 63-81. Raven Press, New York
90. Kaufmann, J. and Smale, S. T. (1994) Direct recognition of initiator elements by a component of the transcription factor IID complex. *Genes Dev.* **8**, 821-829
91. Orphanides, G., Lagrange, T. and Reinberg, D. (1996) The general transcription factors of RNA polymerase II. *Genes Dev.* **10**, 2657-83. Review
92. Hernandez, N. (1993) TBP, a universal eukaryotic transcription factor? *Genes Dev.* **7**, 1291-308. Review
93. Burley, S. K. and Roeder, R. G. (1996) Biochemistry and structural biology of transcription factor IID (TFIID). *Annu Rev Biochem.* **65**, 769-99, Review
94. Smale, S. T. (1997) Transcription initiation from TATA-less promoters within eukaryotic protein-coding genes. *Biochim Biophys Acta.* **1351**, 73-88, Review
95. Mueller-Sturm, H. P., Sogo, J. M. and Schaffner, W. (1989) An enhancer stimulates transcription in trans when attached to the promoter via a protein bridge. *Cell*, **58**, 767-77. Erratum in: *Cell* 1989 **59**, 405
96. Britten, R. J. and Davidson, E. H. (1969) Gene regulation for higher cells: a theory. *Science*, **165**, 349-57
97. Johnson, P. F. and McKnight, S. L. (1989) Eukaryotic transcriptional regulatory proteins. *Annu Rev Biochem.* **58**, 799-839. Review
98. Triezenberg, S. J. (1995) Structure and function of transcriptional activation domains. *Curr Opin Genet Dev.* **5**, 190-6. Review

99. Pabo, C. O. and Sauer, R. T. (1992) Transcription factors: structural families and principles of DNA recognition. *Annu Rev Biochem.* **61**, 1053-95. Review
100. Munshi, N., Merika, M., Yie, J., Senger, K., Chen, G. and Thanos, D. (1998) Acetylation of HMG I(Y) by CBP turns off IFN beta expression by disrupting the enhanceosome. *Mol Cell.* **2**, 457-67
101. Benezra, R., Davis, R. L., Lassar, A., Tapscott, S., Thayer, M., Lockshon, D. and Weintraub, H. (1990) Id: a negative regulator of helix-loop-helix DNA binding proteins. Control of terminal myogenic differentiation. *Ann N Y Acad Sci.* **599**, 1-11
102. Momand, J., Zambetti, G. P., Olson, D. C., George, D. and Levine, A. J. (1992) The mdm-2 oncogene product forms a complex with the p53 protein and inhibits p53-mediated transactivation. *Cell.* **69**, 1237-45
103. Vermaak, D. and Wolffe, A. P. (1998) Chromatin and chromosomal controls in development. *Dev Genet.* **22**, 1-6. Review
104. Bustin, M. (2001) Revised nomenclature for high mobility group (HMG) chromosomal proteins. *Trends Biochem Sci.* **26**, 152-153
105. Knapp, S., Muller, S., Digilio, G., Bonaldi, T., Bianchi, M. E. and Musco, G. (2004) The long acidic tail of high mobility group box 1 (HMGB1) protein forms an extended and flexible structure that interacts with specific residues within and between the HMG boxes. *Biochemistry*, **43**, 11992-11997
106. Thomas, J. O. and Travers, A. A. (2001) HMG1 and 2, and related 'architectural' DNA-binding proteins. *Trends Biochem Sci.* **26**, 167-174
107. Agresti, A. and Bianchi, M. E. (2003) HMGB proteins and gene expression. *Curr Opin Genet Dev*, **13**, 170-178
108. Bonaldi, T., Langst, G., Strohner, R., Becker, P. B. and Bianchi, M. E. (2002) The DNA chaperone HMGB1 facilitates ACF/CHRAC-dependent nucleosome sliding. *EMBO J*, **21**, 6865-6873.
109. Travers, A. A. (2003) Priming the nucleosome: a role for HMGB proteins? *EMBO Rep*, **4**, 131-6
110. Wu, C. and Travers, A. (2004) A 'one-pot' assay for the accessibility of DNA in a nucleosome core particle. *Nucleic Acids Res*, **32**, e122
111. Das, D. and Scovell, W. M. (2001) The binding interaction of HMG-1 with the TATA-binding protein/TATA complex. *J Biol Chem*, **276**, 32597-32605
112. Dasgupta, A. and Scovell, W. M. (2003) TFIIA abrogates the effects of inhibition by HMGB1 but not E1A during the early stages of assembly of the transcriptional PIC. *Biochim. Biophys Acta*, **1627**, 101-110



113. van de Wetering, M., Oosterwegel, M., Dooijes, D. and Clevers, H. (1991) Identification and cloning of TCF-1, a T lymphocyte-specific transcription factor containing a sequence-specific HMG box. *EMBO J.* **10**, 123-32
114. Harley, V. R., Jackson, D. I., Hextall, P. J., Hawkins, J. R., Berkovitz, G. D., Sockanathan, S., Lovell-Badge, R. and Goodfellow, P. N. (1992) DNA binding activity of recombinant SRY from normal males and XY females. *Science*, **255**, 453-6
115. Denny, P., Swift, S., Connor, F. and Ashworth, A. (1992) An SRY-related gene expressed during spermatogenesis in the mouse encodes a sequence-specific DNA-binding protein. *EMBO J.* **11**, 3705-12
116. Wegner, M. (1999) From head to toes: the multiple facets of Sox proteins. *Nucleic Acids Res.* **27**, 1409-20. Review
117. Werner, M. H., Huth, J. R., Gronenborn, A. M. and Clore, G. M. (1995) Molecular basis of human 46X,Y sex reversal revealed from the three-dimensional solution structure of the human SRY-DNA complex. *Cell*, **81**, 705-14
118. Giese, K., Cox, J. and Grosschedl, R. (1992) The HMG domain of lymphoid enhancer factor 1 bends DNA and facilitates assembly of functional nucleoprotein structures. *Cell*, **69**, 185-95
119. Koopman, P. (2001) Sry, Sox9 and mammalian sex determination. *EXS.* **91**, 25-56. Review
120. Harley, V. R., Lovell-Gadge, R., Goodfellow, P. N. (1994) Definition of a consensus DNA binding site for SRY. *Nucleic Acids Res.* **22**, 1500-1501.
121. Mertin, S., McDowall, S. G., Harley, V. R. (1999) The DNA binding specificity of Sox9 and other Sox proteins. *Nucleic Acids Res.* **27**, 1359-1364.
122. Kamachi, Y., Uchikawa, M. and Kondoh, H. (2000) Pairing SOX off with partners in the regulation of embryonic development. *Trends Genet.* **16**, 182-187
123. Zhou, G., Lefebvre, V., Zhang, Z., Eberspaecher, H. and de Crombrughe, B. (1998) Three high mobility group-like sequences within a 48-base pair enhancer of the Col2a1 gene are required for cartilage-specific expression in vivo. *J Biol Chem.* **273**, 14989-97
124. De Santa Barbara, P., Bonneaud, N., Boizet, B., Desclozeaux, M., Moniot, B., Sudbeck, P., Scherer, G., Poulat, F. and Berta, P. (1998) Direct interaction of SRY-related protein SOX9 and steroidogenic factor 1 regulates transcription of the human anti-Mullerian hormone gene. *Mol. Cell. Biol.* **11**, 6653-6655
125. Huang, W., Lu, N., Eberspaecher, H. and de Crombrughe B. (2002) A new long form of c-Maf cooperates with Sox9 to activate the type II collagen gene. *J. Biol. Chem.* **277**, 50668-50675
126. Marshall, O. J. and Harley, V. R. (2001) Identification of an interaction between SOX9 and HSP70. *FEBS Lett.* **496**, 75-80
127. Ambrosetti, D. C., Scholer, H. R., Dailey, L. and Basilico, C. (2000) Modulation of the activity of multiple transcriptional activation domains by the DNA binding domains mediates the

- synergistic action of Sox2 and Oct-3 on the fibroblast growth factor-4 enhancer. *J Biol Chem.* **275**, 23387-97
128. Kuhlbrodt, K., Herbarth, B., Sock, E., Hermans-Borgmeyer, I. and Wegner, M. (1998) Sox10, a novel transcriptional modulator in glial cells. *J Neurosci.* **18**, 237-50
129. Hudson, C., Clements, D., Friday, R. V., Stott, D. and Woodland, H. R. (1997) Xsox17alpha and -beta mediate endoderm formation in *Xenopus*. *Cell*, **91**, 397-405
130. Pennisi, D., Bowles, J., Nagy, A., Muscat, G. and Koopman, P. (2000) Mice null for sox18 are viable and display a mild coat defect. *Mol Cell Biol.* **20**, 9331-9336
131. Lefebvre, V., Behringer, R. R. and de Crombrughe, B. (2001) L-Sox5, Sox6 and Sox9 control essential steps of the chondrocyte differentiation pathway. *Osteoarthritis Cartilage*, **9**, S69-75
132. Tsumaki, N., Kimura, T., Matsui, Y., Nakata, K. and Ochi, T. (1996) Separable cis-regulatory elements that contribute to tissue- and site-specific alpha 2(XI) collagen gene expression in the embryonic mouse cartilage. *J Cell Biol* **134**, 1573-1582
133. Bridgewater, L. C., Lefebvre, V. and de Crombrughe, B. (1998) Chondrocyte-specific enhancer elements in the Col11a2 gene resemble the Col2a1 tissue-specific enhancer. *J Biol Chem* **273**, 14998-15006
134. Liu, Y., Li, H., Tanaka, K., Tsumaki, N and Yamada, Y. (2000) Identification of an enhancer sequence within the first intron required for cartilage-specific transcription of the alpha2(XI) collagen gene. *J Biol Chem* **275**, 12712-12718
135. Zhang, P., Jimenez, S. A. and Stokes, D. G. (2003) Regulation of human COL9A1 gene expression. Activation of the proximal promoter region by SOX9. *J Biol Chem* **278**, 117-123
136. Sekiya, I., Tsuji, K., Koopman, P., Watanabe, H., Yamada, Y., Shinomiya, K., Nifuji, A. and Noda, M. (2000) SOX9 enhances aggrecan gene promoter/enhancer activity and is up-regulated by retinoic acid in a cartilage-derived cell line, TC6. *J Biol Chem* **275**, 10738-10744
137. Xie, W. F., Zhang, X., Sakano, S., Lefebvre, V. and Sandell, L. J. (1999) Trans-activation of the mouse cartilage-derived retinoic acid-sensitive protein gene by Sox9. *J Bone Miner.Res* **14**, 757-763
138. Kiss, I., Bösze, Z., Szabó, P., Rencendorj, A., Barta, E. and Deák, F. (1990) Identification of positive and negative regulatory regions controlling expression of the cartilage matrix protein gene. *Mol. Cell. Biol.* **10**, 2432-36
139. Szabó, P., Moitra, J., Rencendorj, A., Rákhely, G., Rauch, T. and Kiss, I. (1995) Identification of a nuclear factor-I family protein-binding site in the silencer region of the cartilage matrix protein gene. *J. Biol. Chem.* **270**, 10212-21
140. Karcagi, I., Rauch, T., Hiripi, L., Rentsendorj, O., Nagy, A., Bösze, Zs. and Kiss, I. (2004) Functional analysis of the regulatory regions of the matrilin-1 gene in transgenic mice reveals modular arrangement of tissue-specific control elements. *Matrix Biol.* **22**, 605-618

141. Hammond, M. P. and Birney, E. (2004) Genome information resources – developments at Ensembl. *Trends Genet.* **20**, 268-272
142. Kulikova, T., Aldebert, P. et al. (2004) The EMBL Nucleotide Sequence Database. *Nucleic Acids Res.* **32** Database issue, D27-30
143. Rice, P., Longden, I. and Bleasby, A. (2000) EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet* **16**, 276-277
144. Morgenstem, B. (1999) DIALIGN 2: improvement of the segment-to-segment approach to multiple sequence alignment. *Bioinformatics* **15**, 211-218
145. Bailey, T. L. and Elkan, C. (1995) The value of prior knowledge in discovering motifs with MEME. *Proc. Int. Conf. Instell. Syst. Mol. Biol.* **3**, 21-29
146. Pfeifer, G. P., Chen, H. H., Komura, J. and Riggs, A. D. (1999) Chromatin structure analysis by ligation-mediated and terminal transferase-mediated polymerase chain reaction. *Methods Enzymol.* **304**, 548-571
147. Maxam, A. and Gilbert, W. (1977) A new method of sequencing DNA. *Proceedings of the National Academy of Sciences, USA* **74**, 560-4
148. Hames, B. D. and Higgins, S. J. (1993) *Gene transcription: A practical approach*. The practical approach series, Oxford University Press
149. Mertin, S., McDowall, G. S. and Harley, V. R. (1999) The DNA binding specificity of SOX9 and other SOX proteins. *Nucleic Acids Res.* **27**, 1359-1364
150. Bridgewater, L. C., Walker, M. D., Miller G. C., Ellison T. A., Holsinger, L. D., Potter, J. L., Jackson, T. L., Chen, R. K., Winkel, V. L., Zhang, Z., McKinney, S. and de Crombrughe, B. (2003) Adjacent DNA sequences modulate Sox9 transcriptional activation at paired Sox sites in three chondrocyte-specific enhancer elements. *Nucleic Acids Res.* **31**, 1541-1553
151. Pöhler, J. R., Norman, D. G., Bramhan, J., Bianchi M. E., Lilley, D. M., (1998) HMG box proteins bind to four-way DNA junctions in their open conformation. *The EMBO Journal.* **17**, 817-826.
152. Rentsendorj, O., Nagy, A., Sinko, I., Daraba, A., Barta, E. and Kiss, I. (2005) Highly conserved proximal promoter element harboring paired Sox9-binding Sites contributes to the tissue- and developmental stage-specific activity of the matrilin-1 gene. *Biochem J.* **389**, 705-16
153. Uchikawa, M., Ishida, Y., Takemoto, T., Kamachi, Y. and Kondoh, H. (2003) Functional analysis of chicken Sox2 enhancers highlights an array of diverse regulatory elements that are conserved in mammals. *Dev. Cell* **4**, 509-519
154. Kamachi, Y., Uchikawa, M., Collignon, J., Lovell-Badge, R. and Kondoh, H. (1998) Involvement of Sox1, 2 and 3 in the early and subsequent molecular events of lens induction. *Development*, **125**, 2521-32

155. Nishiguchi, S., Wood, H., Kondoh, H., Lovell-Badge, R. and Episkopou, V. (1998) Sox1 directly regulates the  $\alpha$ -crystallin genes and is essential for lens development in mice. *Genes Dev.* **12**, 776-781
156. Furlong, E. E., Rein, T. and Martin, F. (1996) YY1 and NF1 both activate the human p53 promoter by alternatively binding to a composite element, and YY1 and E1A cooperate to amplify p53 promoter activity. *Mol. Cell. Biol.* **16**, 5933-5945
157. Niehof, M., Kubicka, S., Zender, L., Manns, M. P. and Trautwein, C. (2001) Autoregulation enables different pathways to control CCAAT/enhancer binding protein beta (C/EBP beta) transcription. *J. Mol. Biol.* **309**, 855-868
158. Ducy, P. and Karsenty, G. (1995) Two distinct osteoblast-specific cis-acting elements control expression of a mouse osteocalcin gene. *Mol. Cell. Biol.* **15**, 1858-1869
159. Zhou, G., Garofalo, S., Mukhopadhyay, K., Lefebvre, V., Smith, C. N., Eberspaecher, H. and de Crombrughe, B. (1995) A 182 bp fragment of the mouse pro $\alpha$ 1(II) collagen gene is sufficient to direct chondrocyte expression in transgenic mice. *J. Cell Sci.* **108**, 3677-3684
160. Xie, W. F., Zhang, X. and Sandell, L. J. (2000) The 2.2-kb promoter of cartilage-derived retinoic acid-sensitive protein controls gene expression in cartilage and embryonic mammary buds of transgenic mice. *Matrix Biol.* **19**, 501-509
161. Sock, E., Pagon, R. A., Keymolen, K., Lissens, W., Wegner, M. and Scherer, G. (2003) Loss of DNA-dependent dimerization of the transcription factor SOX9 as a cause for campomelic dysplasia. *Hum. Mol. Genet.* **12**, 1439-47
162. Peirano, R. I. and Wegner, M. (2000) The glial transcription factor Sox10 binds to DNA both as monomer and dimer with different functional consequences. *Nucleic Acids Res.* **28**, 3047-55
163. Bernard, P., Tang, P., Liu, S., Dewing, P., Harley, V. R. and Vilain, E. (2003) Dimerization of SOX9 is required for chondrogenesis, but not for sex determination. *Hum. Mol. Genet.* **12**, 1755-65
164. Ge, H., and Roeder, R. G. (1994) The high mobility group protein HMG1 can reversibly inhibit class II gene transcription by interaction with the TATA-binding protein. *J. Biol. Chem.* **269**, 17136-40
165. Oñate, S. A., Prendergast, P., Wagner, J. P., Nissen, M., Reeves, R., Pettijohn, D. E., and Edwards, D. P. (1994) The DNA-bending protein HMG-1 enhances progesterone receptor binding to its target DNA sequences. *Mol. Cell. Biol.* **14**, 3376-3391
166. Sutrias-Grau, M., Bianchi, M. E. and Bernues, J. (1999) High mobility group protein 1 interacts specifically with the core domain of human TATA box-binding protein and interferes with transcription factor IIB within the pre-initiation complex. *J Biol Chem.* **274**, 1628-34
167. Murray, D., Precht, P., Balakir, R. and Horton, Jr, W. E. (2000) The transcription factor deltaEF1 is inversely expressed with type II collagen mRNA and can repress Col2a1 promoter activity in transfected chondrocytes. *J Biol Chem* **275**, 3610-3618

168. Takagi, T., Moribe, H., Kondoh, H. and Higashi, Y. (1998) DeltaEF1, a zinc finger and homeodomain transcription factor, is required for skeleton patterning in multiple lineages. *Development*, **125**, 21-31
169. Seki, K., Fujimori, T., Savagner, P., Hata, A., Aikawa, T., Ogata, N., Nabeshima, Y. and Kaechoong, L. (2003) Mouse Snail family transcription repressors regulate chondrocyte extracellular matrix, type II collagen and aggrecan. *J Biol Chem*. **278**, 41862-70
170. Krebsbach, P. H. and Yamada, Y. (1996) Role of the first intron in the regulation of the COL1A1 gene. *Matrix Biol*. **15**, 64
171. Tanaka, K., Matsumoto, Y., Nakatani, F., Iwamoto, Y. and Yamada, Y. (2000) A zinc finger transcription factor, alphaA-crystallin binding protein 1, is a negative regulator of the chondrocyte-specific enhancer of the alpha1(II) collagen gene. *Mol Cell Biol* **20**, 4428-4435
172. Tsumaki, N., Kimura, T., Tanaka, K., Kimura, J. H., Ochi, T. and Yamada, Y. (1998) Modular arrangement of cartilage- and neural tissue-specific cis-elements in the mouse alpha2(XI) collagen promoter. *J Biol Chem* **273**, 22861-64
173. Huang, Z., Xu, H. and Sandell, L. (2004) Negative regulation of chondrocyte differentiation by transcription factor AP-2alpha. *J Bone Miner Res* **19**, 245-255

## SUMMARY OF THE THESIS

### Introduction

Cartilage anlagen develop in the embryo before bone, and thus provide the first skeleton of the embryo. It acts as the main source of body longitudinal growth that provides structural templates and induction signals for the formation of most bones through a process called endochondral ossification (Olsen et al. 2000). Cartilage is an elaborate network of large macromolecules, synthesised and deposited by the specialized cells, called chondrocytes. They deposit these macromolecules around themselves to form an extracellular matrix (ECM).

Chondrocytes express a number of specific genes, such as types II, VI, IX, and XI collagens, aggrecan, link protein, CD-Rap, COMP, and matrilin-1. It is clear that components of the cartilage matrix itself play important roles in either modulating or maintaining the phenotype of chondrocytes and their correct organization in the growth plate (Mundlos and Olsen. 1997, Erlebacher et al. 1995). In addition, the precise expression pattern depends on a balance of positive and negative transcription factors, proteins that control the synthesis of mRNAs from these specific genes. In cartilage, a number of studies indicate that Sox9 is an activating transcription factor, indispensable for chondrogenesis. During chondrogenesis, Sox9 is expressed in chondroprogenitors and chondroblasts, and activates the cartilage-specific genes, such as *Col2a1*, *Col11a2*, *aggrecan*, and *CD-Rap*. The study of chimeric mice that harbour Sox9<sup>-/-</sup> cells revealed that cells without Sox9 do not express chondrocyte marker genes such as *Col2a1*, *Col9a2*, *Col11a2*, and *aggrecan*. Sox9 binds to and directly activates the cartilage-specific regulatory elements of *Col2a1* as well as *Col11a2*, which can direct cartilage-specific expression in transgenic mice. Sox9 also transactivates the promoter activities of *Col9a1*, *aggrecan*, and *CD-Rap*.

Additional Sox factors, such as L-Sox5 and Sox6, are also involved in chondrogenesis (Smits et al. 2001). The expression of L-Sox5 and Sox6 requires Sox9 (Akiyama et al. 2004). They are activated in prechondrocytes and highly expressed in chondroblasts in all developing cartilage elements of the mouse embryo (Lefebvre et al. 1998). Although individual Sox5 or Sox6 null mice are born with minor cartilage defects, double knockout mice develop a severe, generalized

chondrodysplasia. In the double null mice, prechondrocytes are unable to progress along the differentiation pathway, expressing a very low level of early cartilage genes, such as *Col2a1* and *aggrecan*, but they fail to turn on stage-specific matrix genes such as *COMP* and *matrilin-1* (Smits et al. 2001). *In vitro* experiments have suggested that, besides Sox9, L-Sox5 and Sox6 are also involved in *Col2a1* expression. They appear to form a large complex with each other and other nuclear proteins in chondrocytes. Therefore, it is believed that the three Sox cooperatively activate the *Col2a1* (Lefebvre et al. 1997, 1998, Bell et al. 1997). They were also able to suppress expression of markers for hypertrophic chondrocytes, thus confirming that L-Sox5, Sox6, and Sox9 constitute a master chondrogenic trio.

Matrilin-1 is tightly associated with aggrecan (Hauser et al. 1996) and type II collagen fibrils (Winterbottom et al. 1992), the two major macromolecular networks of cartilage. It also forms collagen-independent pericellular filaments via the vWFA domains (Chen et al. 1999), and interacts with  $\alpha 1\beta 1$  integrin (Makihira et al. 1999). Therefore, it seems to perform an adapter function in the assembly of the cartilaginous matrix. Among cartilage protein genes, the matrilin-1 has a unique expression pattern, restricted to certain chondrocyte developmental stages and distinct zones of the growth plate. It has been confirmed that the *matrilin-1* gene is activated only in late proliferative chondrocytes *in vitro* (Muratoglu et al. 1995, Szuts et al. 1998). This suggests that the regulation of the gene may involve similar as well as different molecular mechanisms as compared to other cartilage protein genes, such as *Col2a1*. To get insight into the common and distinct molecular mechanisms controlling cartilage-specific gene expression, we aimed to study DNA elements and transcription factors, involved in the transcriptional regulation of the matrilin-1 gene.

### **Aims of study**

Functional analysis of the main regulatory regions of the gene revealed that the long promoter with and without the intronic enhancer region, as well as the short promoter with the intronic enhancer region equally direct the transgene expression to distinct zones of the growth plate in transgenic mice (Karcagi et al. 2004). Based on this observation, we raised the question, whether the short promoter

may have a role in the tissue-specific regulation and if it involves cartilage-specific and/or developmental stage-specific control elements. The following specific aims were set out:

- to delineate *cis*-regulatory elements that are important for the transcriptional regulation of the matrilin-1 gene and to identify proteins binding to these sites, especially focusing on those sites, which possibly function via interacting with Sox-family proteins in the short promoter region.
- to verify Sox protein-binding sites in the short promoter using *in vitro* gelshift and supershift assays.
- to reveal tissue-specific binding of transcription factors to the short promoter region using *in vivo* footprinting.
- to study the role of these putative elements in the transcriptional regulation of the matrilin-1 gene.

## Methods

In order to accomplish our aims we performed the following experimental methods:

- ✓ Sequence analysis
- ✓ Isolation of genomic DNA from tissue culture
- ✓ *In vivo* footprinting with dimethyl sulphate and UV light
- ✓ Preparation of cell extracts for EMSA
- ✓ Synthesis and purification of bacterially expressed GST-SOX9
- ✓ EMSA and supershift experiments
- ✓ Luciferase reporter gene constructs for transient expression assay
- ✓ Site directed mutagenesis
- ✓ Transient expression assay

## Results and Discussion

In this study, we have shown that the short promoter of chicken matrilin-1 gene contains at least two *cis*-acting Sox-specific islands labelled as Pe1 and Ine. Both elements interact with chondrogenic transcription factors of Sox9, L-Sox5 and Sox6 *in vitro*. Point mutations in the Sox



motifs of either element interfered with or altered the formation of nucleoprotein complexes *in vitro* and significantly decreased the reporter gene activity in CEC in transient expression assays, stressing the importance of these elements in the transcriptional upregulation of the matrilin-1 gene. Transient expression studies confirmed that both Pe1 and Ine elements significantly contribute to the moderate activity of the short promoter in chondrocytes. Based on the data, we hypothesize that the element does not drive the high cartilage-specific expression of the promoter as an enhancer, but may rather act by modulating the promoter activity and mediating the effect of distal promoter and intronic enhancer elements. In addition, we propose that distal promoter and intronic elements may also function by forming multi-protein complexes via interacting with Pe1 and Ine.

We were able to show *in vivo* occupancy of the Sox motifs in genomic footprinting in the expressing cell type, but not in the non-expressing, which support the involvement of Pe1 and Ine in the tissue-specific regulation of the gene. Apart from Pe1 and Ine, in chondrocytes, footprints were also observed at the NFI contact points of the SI and SII elements that were identified previously (Szabó et al. 1995). As opposed to this, the complete absence of footprints in the short promoter in fibroblasts in repeated experiments indicated no factors bound to their recognition sequences in the non-expressing cell type. This suggests that regulation at the chromatin level can be involved in the activation of the gene in chondrocytes.

Although Pe1 can bind to Sox9, L-Sox5 and Sox6 *in vitro*, it may show a preference for Sox9 *in vivo*, as the inverted Sox motifs carried by Pe1 are more similar to the preferred binding sites of Sox9, than to those of L-Sox5/Sox6. In our experiments, GST–Sox9 formed only a single complex with Pe1 and Ine. Even though Sox9 is not capable of forming homodimers in solution (Lefebvre et al. 1998, Sock et al. 2003, Peirano and Wegner. 2000), consistent with observations from other laboratories (Bridgewater et al. 2003, Bernard et al. 2003), we found that it could bind only to intact inverted pairs of Sox motifs, thus supporting the conclusion that Sox9 dimerization might have occurred upon DNA binding. As opposed to this, Sox9 was reported to bind as a monomer to *cis* elements involved in sex determination (Bernard et al. 2003).

Collateral analysis based on computer searches revealed that Pe1 element of the chicken matrilin-1 gene is highly conserved under evolutionary pressure between chicken and mammals

implies that the element performs a very important function in the transcriptional regulation of the matrilin-1 gene in amniotes.

This is the first report demonstrating that paired Sox-specific sites mapped within Ine, that is close to the TATA motif can play a functional role in the transcriptional activation of a cartilage protein gene. The close proximity of the Sox-binding sites of Ine to the TATA box raises questions regarding the possible interactions of Sox proteins with general transcription machinery. An involvement of HMG1 domain proteins in the regulation of gene transcription has been suspected for a long time. It has been reported for example that the abundant HMG1 protein, which binds angled structures in the DNA without any sequence specificity, can interact in a species-specific manner with the core domain of the TATA box-binding protein (TBP), and block the formation of the preinitiation complex (PIC) by preventing the binding of TFIIB to TBP (Ge and Roeder. 1994, Oñate et al. 1994). Differing from the HMG1 protein, Sox9 binds to the DNA in a sequence-specific manner and has a transactivation domain. We hypothesize that, according to the intrinsic flexibility of DNA bending and cooperative binding ability with multiple transcription factors, Sox proteins (L-Sox5, Sox6 and Sox9) may interact with the components of the general transcription machinery assembled on Ine, which may generate a surface complementary to the RNA polymerase II transcription machinery. It is yet to be confirmed that this interaction or combination of general transcription machinery really function on the natural Ine element of the matrilin-1 gene.

By bending the DNA, HMG-box proteins are known to promote the binding of other transcription factors to the DNA. Lining up with these observations, our results suggest that, in addition to Sox proteins, other transcription factors may also be involved in the activity of the short matrilin-1 promoter, thereby contributing to the developmental stage-specific activity of the matrilin-1 gene. Therefore, future studies will be needed for mapping the putative cofactor-binding sites and identifying the interacting Sox partner factors on the Pe1 and Ine elements.

These findings imply for the first time an important role of cis-regulatory elements that are functioning via Sox family proteins in the transcriptional upregulation of the chicken matrilin-1 gene.

## List of novel findings

1. Throughout 1.2-kb promoter region tested between positions -1137 and +64, we identified two potential Sox-specific islands labelled as Pe1 and Ine in the short promoter of chicken matrilin-1 gene.
  - a. Pe1 element includes a pair of inverted motifs highly similar to the AGAACAATGG motif, which was shown to be the preferred binding site of Sox9 *in vitro*.
  - b. Ine1 harbors two pairs of inverted putative Sox motifs with 5/10, 6/10, 7/10 and 4/9 identity with the consensus Sox-binding sites CA/TTTGA/TA/T *in vitro*.
2. We observed *in vivo* occupancy of the Sox motifs in genomic footprinting in the expressing cell type, but not in the non-expressing, which support the involvement of Pe1 and Ine in the tissue-specific regulation of the gene. Apart from Pe1 and Ine, in chondrocytes, footprints were also visible at the NFI contact points of the SI and SII elements that were identified previously by Szabó et al in 1995.
3. We provided evidence that both elements (Pe1 and Ine) can interact with chondrogenic transcription factors of Sox9, L-Sox5 and Sox6 *in vitro*.

Eventhough, Pe1 contain pair of inverted Sox9 motifs, we noticed that the element form only a single nucleoprotein complex with GST–Sox9. Point mutations in either Sox motifs interfered with this complex, suggesting that GST-Sox9 could bind only to intact inverted pairs of Sox motifs and that Sox9 dimerization might have occurred upon DNA binding.
4. Using transient expression studies, we confirmed that both Pe1 and Ine elements significantly contribute to the moderate activity of the short promoter in chondrocytes. We hypothesize that the element does not drive the high cartilage-specific expression of the promoter as an enhancer, but may rather act by modulating the promoter activity and mediating the effect of distal promoter and intronic enhancer elements.
5. This is the first report demonstrating that paired Sox-specific sites mapped within Ine, that is close to the TATA motif can play a functional role in the transcriptional activation of a cartilage protein gene. We hypothesize that, according to the intrinsic flexibility of DNA bending and cooperative

binding ability with multiple transcription factors, Sox proteins (L-Sox5, Sox6 and Sox9) may interact with the components of the general transcription machinery assembled on Ine, which may generate a surface complementary to the RNA polymerase II transcription machinery. In addition, distal promoter and intronic elements may also function by forming multi-protein complexes via interacting with Pe1 and Ine.

These findings imply for the first time an important role of cis-regulatory elements that are functioning via Sox family proteins in the transcriptional upregulation of the chicken matrilin-1 gene.

## ÖSSZEFOGLALÁS

### Bevezetés

Az embrionális fejlődés során a csontos váz egy átmeneti porcos vázon át egy igen összetett, többlépcsős differenciálódási folyamat eredményeként alakul ki (Olsen és mti, 2000). A porc, a chondrocyták által termelt makromolekulák bonyolult hálózataként jelenik meg. A chondrocyták által termelt molekulák a sejtek körül extracelluláris mátrixba (ECM) szerveződnek.

A chondrocyták számos porcspecifikus gént fejdenek ki, mint a II., VI., IX. és XI. típusú kollagént, link proteint, CD-Rap, COMP és a matrilin-1 gént. A porcos mátrix fontos szerepet tölt be a chondrocyták fenotípusának meghatározásában és a porcsejtek elrendeződésében a növekedési porckorong különböző övezeteiben (Mundlos és Olsen, 1997; Erlebacher és mti, 1995). A porcmátrix gének megfelelő zonális expresszióját pozitív és negatív hatással bíró transzkripciós faktorok egyensúlya biztosítja. Ezek a faktorok szabályozzák a porc-specifikus gének szintézisét mRNS szinten. Számos kutatási eredmény alapján bebizonyosodott, hogy a Sox9 nélkülözhetetlen fő pozitív transzkripciós faktor a porcban. A porcfejlődés során a Sox9-et a chondroprogenitorok, a chondrocyták termelik. A Sox9 szabályozza a porc-specifikus gének, mint pl. a Col2a1, Col11a2, aggregán és CD-Rap kifejeződését, ugyanis ezek a porc-specifikus gének nem fejdenek ki a Sox9-/- hiányos kiméra egerekben.

A Sox család más tagjai, mint az L-Sox5 és Sox6 is szerepet játszanak a chondrogenézisben (Smits és mti., 2001). A Sox9 szükséges faktor az L-Sox5 és Sox6 kifejeződéséhez (Akiyama és mti., 2004). A L-Sox5 és Sox6 már a prechondroblastokban aktiválódik és fokozott expressziót mutat a fejlődő porc chondroblastjaiban (Lefebvre és mti., 1998). Habár az L-Sox5 és Sox6 hiányos egerek kis mértékű elváltozást mutatnak a fejlődő porcban, a Sox5/Sox6 kettős transzgenikus egerek általános chondrodysplasiában szenvednek. Ezekben az egerekben a prechondroblastok nem képesek differenciálódni, kis mértékben fejdenek ki a korai porc-specifikus géneket, mint például a Col2a1-t és aggregánt, ill. nem képesek a zonálisan kifejeződő mátrix gének, mint a COMP és matrilin-1 termelésére (Smits és mti., 2001). In vitro kísérletek kimutatták, hogy a Sox9 mellett az L-Sox5 és a Sox6 is szerepet játszik a Col2a1 expressziójában (Lefebvre és mti., 1997; 1998; Bell és mti., 1997).

Feltehetőleg a három Sox faktor együttesen serkenti a Col2a1 expresszióját, továbbá képes gátolni a hipertóf porcsejtek génexpresszióját. Chondrocytákban az L-Sox5, Sox6, Sox9 együttesen és más nukleáris fehérjékkel kölcsönhatva a chondrocytákban szabályozó komplexekbe rendeződnek. Mindezek az eredmények arra utalnak, hogy a három Sox fehérje a chondrogenesis fő transzkripciós faktorai közé sorolható.

A matrilin-1 szorosan kapcsolódik a porcós mátrix fő molekuláihoz, az aggrecánhoz (Hauser és mti., 1996) és a II. típusú kollagénhez (Winterbottom és mti., 1992). Ugyanakkor a vWFA doméneken keresztül kollagén rostoktól független pericelluláris filamentumok képzésére is alkalmas (Chen és mti., 1999) és kapcsolódik az  $\alpha 1\beta 1$  integrinhez is (Makihira és mti., 1999). Ennek alapján elmondható, hogy bizonyos adapter funkciót is betölt a porcós mátrix szerveződésében. A porcfehérjék génjei közül a matrilin-1 egyedi expressziós mintázatot mutat. A növekedési porckorongban zonálisan fejeződik ki. In vitro kísérletekben bebizonyították, hogy a matrilin-1 csak a késői proliferatív chondrocytákban kapcsol be így ennek a stádiumnak a jellegzetes marker génjeként értékelhető (Muratoglu és mti., 1995; Szuts és mti., 1998). A matrilin-1 gén szabályozása bizonyos mértékben feltehetőleg megegyezik más porc mátrix gének szabályozásával de, különbséget is mutat azokhoz képest. A jelen dolgozatban a matrilin-1 transzkripciós szabályozásában szerepet játszó DNS elemek, és transzkripciós faktorok meghatározását tűztük ki célul.

### **Célkitűzések**

A matrilin-1 fő szabályozó régiójának funkcionális vizsgálata kimutatta, hogy a hosszú és a rövid promoter az introni enhancer régióval ill. ez utóbbi nélkül, a növekedési porckorong meghatározott zónájába irányítja a transzgén kifejeződését a vizsgált transzgenikus egerekben (Karcagi és mti., 2004). Mindezek ismeretében felvetődik a kérdés, hogy vajon a rövid promoterek szerepe lehet-e a szövet-specifikusság szabályozásában ill., hogy ez a szabályzó mechanizmus milyen porc és/vagy fejlődés-specifikus szabályzó elemek jelenlétét feltételezi.

- A matrilin-1 gén transzkripciós szabályozásában szerepet játszó cisz-szabályozó elemek meghatározása a rövid promoter régiójában és azoknak a fehérjéknek az azonosítása, melyek a Sox fehérjékkel való kölcsönhatásuk révén kötődnek ezekhez a helyekhez.

- A Sox-kötőhelyek jelenlétének igazolása a rövid promoteren belül in vitro gelshift és supershift kísérletek alkalmazásával.
- Transzkripciós faktorok szövet-specifikus kötődésének igazolása a rövid promoterhez in vivo footprint technikával.
- Az egyes DNS elemeknek a matrilin-1 transzkripciós szabályozásában betöltött szerepének tanulmányozása.

### **Alkalmazott módszerek**

- ✓ Szekvencia analízis
- ✓ Genomi DNS izolálása sejt kultúrákból
- ✓ In vivo footprint dimetil-szulfát és UV besugárzás alkalmazásával
- ✓ Sejt kivonat készítés az EMSA kísérletekhez
- ✓ Baktériumban expresszált GST-SOX9 tisztítása
- ✓ EMSA és supershift
- ✓ Luciferáz riportergén konstrukciók létrehozása
- ✓ Tranziens expressziós kísérletek
- ✓ Célzott mutagenesis

### **Az eredmények ismertetése és megvitatása**

Kimutattuk, hogy a csirke matrilin-1 gén rövid promotere legalább két olyan DNS elemet tartalmaz, melyekben Sox specifikus kötőhelyek találhatóak. Ez a két cisz-szabályozó régió a Pe1 és az Ine jelölést kapta in vitro körülmények között mindkét elem kölcsönhatásba lép a Sox9, L-Sox5 és Sox6 chondrogén transzkripciós faktorokkal. A Pe1 és Ine elemek Sox kötőhelyeinek pontmutációja in vitro gelshift kísérletekben megváltoztatta a kialakuló nukleoprotein komplex mintázatát és csirke primer chondrocyta sejt kultúrában végzett tranziens expressziós kísérletekben szignifikánsan csökkentette a riportergén aktivitását. A kísérleti eredmények arra utalnak, hogy a Pe1 és Ine elem lényegesen befolyásolja a matrilin-1 gén promoterének az aktivitását azáltal, hogy közvetíti a disztális promoter és az introni enhancer elemek hatását. Hipotézisünk szerint a disztális promoter elemeken és

az introni elemeken kialakuló fehérje komplexek a Pe1 és az Ine elem közvetítésével multiprotein komplexek kialakításában vesznek részt.

A matrilin-1-et kifejező sejtekben genomi footprint kísérletekben *in vivo* kötődést mutattunk ki a Pe1 és Ine elem Sox motívumainál. Ezt a kötődést nem sikerült kimutatnunk a matrilin-1-et nem expresszáló sejtekben. Ezek az adatok arra utalnak, hogy a Pe1 és az Ine szerepet játszanak a matrilin-1 szövet-specifikus szabályozásában. Chondrocytákban a Pe1 és Ine elemek footprintjei mellett footprinteket azonosítottunk a már leírt SI és SII elemek (Szabó és mti., 1995) NF1 transzkripciósfaktort kötő régióiban is. A negatív kontrollként használt, matrilin-1-et nem expresszáló fibroblasztokban nem sikerült footprintet azonosítanunk, ami arra utal, hogy a kromatin szinten történő szabályozás szerepet játszhat a gének aktiválásában.

A Sox9, L-Sox5, Sox6 *in vitro* kapcsolódik a Pe1 elemhez. *In vivo* kísérletekben viszont a Sox9 mutat nagyobb kötési affinitást, ugyanis a Pe1 elemben található fordított Sox motívum nagyobb hasonlóságot mutat a Sox9 által preferált kötőhelyhez, mint az L-Sox5/Sox6 kötőhelyekhez képest. Kísérleteinkben a GST-Sox9 csak egy nukleoprotein komplexet alkot a Pe1 és Ine elemekkel. Annak ellenére, hogy oldatban a Sox9 nem képes homodimerek kialakítására (Lefebvre és mti., 1998; Sock és mti., 2003; Peirano és Wegner, 2000) mások eredményeihez hasonlóan (Bridgewater és mti., 2003; Bernard és mti., 2003) mi is azt tapasztaltuk, hogy a Sox9 csak az intakt fordított Sox motívumokhoz kapcsolódik, alátámasztva azt a feltevést miszerint DNS kötés során a Sox9 képes dimerek létrehozására. Ezzel szemben, kimutatták, hogy a Sox9 képes monomerként is kötődni a nem meghatározottságban szerepet játszó cisz-szabályozó elemekhez (Bernard és mti., 2003).

Számítógépes szekvencia elemzésből kiderült, hogy a matrilin-1 gén Pe1 eleme nagymértékben konzerválódott a csirkénél és az emlősöknél, ami arra utal, hogy ez az elem fontos szerepet játszik a matrilin-1 szabályozásában az amniotáknál.

Először mutattuk ki, hogy az Ine elembe azonosított páros Sox-specifikus kötőhelyek a TATA motívumhoz közel helyezkednek el, és funkcionális szerepet játszhatnak a porc gének transzkripcióaktiválásában. A Sox kötőhelyek közelsége a TATA boxhoz felveti annak az esélyét, hogy a Sox fehérjék kölcsönhatásba léphetnek az általános transzkripció apparátussal is. Már rég feltételezik, hogy a HMG1 fehérjék szerepet játszanak a gének transzkripció szabályozásában. Kimutatták, hogy a



HMG1 fehérje, mely különösebb szekvencia-specifikusság nélkül kötődik a hajlított DNS-hez, faj-specifikus módon kölcsön tud hatni a TATA box-kötő fehérjével, így gátolva a TFIIB kötődését, és ezzel megakadályozva a preiniciációs komplex kialakulását (Ge és Roeder, 1994; Oñate és mti., 1994). A HMG1 fehérjével szemben a Sox9 szekvencia specifikus módon kötődik a DNS-hez és rendelkezik egy transzaktivációs doménnel is. Feltételezzük, hogy a Sox fehérjék (L-Sox5, Sox6 és Sox9) a DNS kötő képességük, valamint más transzkripciós faktorokkal való kooperatív kölcsönhatási képességük révén kötődni tudnak az Ine elemen kialakuló transzkripciós apparátushoz és ezzel elősegíthetik az RNS polimeráz II kötődését és a transzkripció elindulását. További kísérletek szükségesek annak igazolására, hogy a Sox faktorok ténylegesen kölcsönhatnak az általános transzkripciós rendszerrel a matrilin-1 gén Ine elemén.

A HMG boxot tartalmazó fehérjék képesek a DNS meghajtására és ezzel más transzkripciós faktorok kötődését segítik elő. Ezzel a megfigyeléssel összhangba a saját kísérleti adataink arra utalnak, hogy a Sox fehérjék mellet más transzkripciós faktorok is szerepet játszhatnak a matrilin-1 gén rövid promoterének aktivitásában, meghatározva ezáltal a matrilin-1 gén fejlődés-specifikus kifejeződését. További kísérletek szükségesek a Pe1 és Ine elemeken feltételezett kofaktor kötőhelyek feltérképezésére és azon transzkripciós faktorok azonosítására, melyek az említett elemeken a Sox fehérjékkel kölcsönhatásba lépnek.

A dolgozatban ismertetett eredmények igazolják a Sox fehérje család tagjainak kötése által ható cisz-szabályozó elemek fontosságát a csirke matrilin-1 gén transzkripciós szabályozásában.

### **A dolgozatban leírt új eredmények**

1. A csirke matrilin-1 gén promoterén 1.2 kb hosszú szakaszon, a -1137 és +64-es régióban két potenciális Sox kötőhelyet találtunk. A két Sox kötő szigetet Pe1 és Ine-vel jelöltük.
  - a. A Pe1 elem tartalmaz egy pár fordított motívumot, mely nagymértékben hasonlít az AGAACAATGG motívumhoz. In vitro adatok szerint a Sox9 nagy affinitással kötődik ehhez a motívumhoz.
  - b. Az Ine elemekben két pár fordított Sox motívum található, mely in vitro 5/10, 6/10, 7/10 és 4/9 hasonlóságot mutat az általános CA/TTTGA/TA/T Sox kötőhellyel.

2. A matrilin-1-et kifejező sejtekben genom footprint kísérletekben in vivo kötődést mutattunk ki a Pe1 és Ine elem Sox motívumainál. Ezt a kötődést nem sikerült kimutatnunk a matrilin-1-et nem expresszáló sejtekben. Ezek az adatok arra utalnak, hogy a Pe1 és az Ine szerepet játszanak a matrilin-1 szövet-specifikus szabályozásában. Chondrocytákban a Pe1 és Ine elemektől függetlenül footprinteket azonosítottunk a már leírt SI és SII elemek (Szabó és mti, 1995) NF1 transzkripció faktort kötő régióiban is.
3. Kimutattuk, hogy mindkét elem (Pe1 és Ine) in vitro kölcsönhat a Sox9, L-Sox5 és Sox6 porc transzkripció faktorokkal.
  - Kísérleteinkben a GST-Sox9 csak egy nukleoprotein komplexet alkot a Pe1 és Ine elemekkel. A Sox9 csak az intakt fordított Sox motívumokhoz kapcsolódik, alátámasztva azt a feltevést miszerint DNS kötés során a Sox9 képes dimerek létrehozására
4. A tranziens expressziós kísérleti eredmények arra utalnak, hogy a Pe1 és Ine elem lényegesen befolyásolja a matrilin-1 gén rövid promotérének az aktivitását a chondrocitákban azáltal, hogy közvetíti a disztális promoter és az introni enhancer elemek hatását.
5. Először mutattuk ki, hogy az Ine elembe azonosított páros Sox-specifikus kötőhelyek a TATA motívumhoz közel helyezkednek el, és funkcionális szerepet játszhatnak a porc gének transzkripció aktiválásában. Feltételezzük, hogy a Sox fehérjék (L-Sox5, Sox6 és Sox9) a DNS kötő képességük, valamint más transzkripció faktorokkal való kooperatív kölcsönhatási képességük révén kötődni tudnak az Ine elem kialakuló transzkripció apparátushoz és ezzel elősegítik az RNS polimeráz II kötődését és a transzkripció elindulását.